TENOR
ZURICH
2024
INTERNATIONAL CONFERENCE
ON TECHNOLOGIES FOR MUSIC
NOTATION & REPRESENTATION

Institute for Computer Music and Sound Technology
Zurich University of the Arts
Switzerland

4–6 April 2024

Ninth International Conference on Technologies for Music Notation and Representation

Edited by Philippe Kocher

TENOR

INTERNATIONAL CONFERENCE
ON TECHNOLOGIES FOR MUSIC
NOTATION & REPRESENTATION

The Ninth International Conference on Technologies for Music Notation and Representation
Zurich, Switzerland 4–6 April 2024

Edited by Philippe Kocher

Proceedings published by Zurich University of the Arts
ISBN 978-0-6481592-0-9
https://doi.org/10.5281/zenodo.13144433

https://www.zhdk.ch/tenor24
https://tenor-conference.org

Preface

The conference team at ICST – Institute for Computer Music and Sound Technology – is glad to present the proceedings of the 9th Tenor Conference 2024, hosted from 4 to 6 April 2024 at Toni Areal in Zurich, the campus of the Zurich University of the Arts.

Among 82 submissions, 48 participants were selected to present 21 papers, 6 posters, 6 workshops and demos, 5 installations and 14 compositions. A Best Paper Award for exceptional research contribution was offered by the Computer Music Journal and curated by our group.

The call for submissions proposed a focus on the creative process, highlighting the role of the sketch in music and sound-related creative practices. Keynotes by Laura Zattra, Svetlana Maras and Philippe Esling addressed the conference focus from three different angles: the philological perspective of a musicologist working with archives, the creative perspective of a composer experimenting with notation in different formats involving technology, and the technological perspective of a computer scientist developing new approaches to artificial creative intelligence.

A great effort was made to offer the best possible conditions for the presentation of sonic works, reflecting the research on the performance of electroacoustic music realised in the last decade at ICST. In addition to sound installations and performances by conference participants, the artistic program included collaborations with the Zurich-based Ensemble Collegium Novum, faculty members and students of the Music Department of the ZHDK, as well as invited soloists.

We endeavoured to make the conference more accessible to early career researchers and local artists without institutional affiliation through offering scholarships, and provided an open live stream of keynotes and paper sessions for other members of the community not present in Zurich. These will be made available as video documentation for future reference.

We thank all researchers and artists responding to our call, all the authors present in these proceedings, our partners Collegium Novum and Computer Music Journal as well as the Tenor Steering Committee and all colleagues at the ZHdK supporting Tenor Zurich 2024 for their commitment and trust. A special thanks goes to Philippe Kocher, who was responsible for the edition of these proceedings.

Germán Toro Pérez
Tenor 2024 conference chair

Organisation

Conference Team

Germán Toro Pérez, conference chair

Leandro Gianini, technical supervision
Rama Gottfried, scientific program co-chair
Philippe Kocher, scientific program co-chair
Céline Lutz, organisation
Hannah Mehler, communication
Leandra Nussbaumer, organisation
Micha Seidenberg, artistic program chair
Sophie Vögele, diversity advisor

Paper Commitee

Miriam Akkermann	TU Dresden, Germany
Alessandro Anatrini	Hochschule für Musik und Theater Hamburg, Germany
Constantin Basica	CCRMA, Stanford University, USA
Jonathan Bell	Aix-Marseille University, France
Lucas Bennett	ICST, Zurich University of the Arts, Switzerland
Jean Bresson	Ableton, Germany
James Tsz-Him Cheung	Hochschule für Musik und Theater Hamburg, Germany
Marko Ciciliani	IEM, University of Music and Performing Arts Graz, Austria
Anthony De Ritis	Northeastern University, Boston, USA
Gerhard Eckel	IEM, University of Music and Performing Arts Graz, Austria
Aaron Einbond	City, University of London, UK
Jason Freeman	Georgia Institute of Technology, USA
Daniele Ghisi	CNMAT, University of California, Berkeley, USA
Jean-Louis Giavitto	IRCAM/CNRS, France
Rama Gottfried	ICST, Zurich University of the Arts, Switzerland
Thomas Grill	University of Music and Performing Arts Vienna, Austria
Karim Haddad	IRCAM, France
Georg Hajdu	Hochschule für Musik und Theater Hamburg, Germany
Richard Hoadley	Independent Artist and Scholar, UK
Cat Hope	Monash University, Australia
Daniel Hug	Zurich University of the Arts, Switzerland
Hanna Järveläinen	ICST, Zurich University of the Arts, Switzerland
David Kim-Boyle	The University of Sydney, Australia
Volkmar Klien	Anton Bruckner Privatuniversität, Austria
Philippe Kocher	ICST, Zurich University of the Arts, Switzerland
Eric Larrieux	ICST, Zurich University of the Arts, Switzerland
Manolo Müller	ICST, Zurich University of the Arts, Switzerland
Teoma Naccarato	Independent Artist and Scholar, Germany
Martin Neukom	ICST, Zurich University of the Arts, Switzerland
Kilian Schwoon	University of the Arts Bremen, Germany
Mattias Sköld	Royal College of Music & Royal Institute of Technology, Sweden
Ryan Smith	State University of New York at Oneonta, USA
Daniel Spreadbury	Steinberg Media Technologies GmbH, Germany

Germán Toro Perez	Zurich University of the Arts, Switzerland
Pierre Alexandre Tremblay	University of Huddersfield, UK
Jeffrey Trevino	California State University, Monterey Bay, USA
Paul Turowski	University of Liverpool, UK
Craig Vear	IOCT, University of Nottingham, UK

Sonic Works / Music Committee

Hannah Mehler	Zurich University of the Arts, Switzerland
Lars Mlekusch	Zurich University of the Arts, Switzerland
Micha Seidenberg	Zurich University of the Arts, Switzerland
Germán Toro Perez	Zurich University of the Arts, Switzerland
Karin Wetzel	Basel Academy of Music, Switzerland

TENOR International Steering Committee

Jean Bresson	Ableton, Germany
Georg Hajdu	Hochschule für Musik und Theater Hamburg, Germany
Richard Hoadley	Independent Artist and Scholar, UK
Cat Hope (General Chair)	Monash University, Australia
Craig Vear	IOCT, University of Nottingham, UK

Table of Contents

THE COMPOSER PHOTOGRAPH: A FRAMEWORK TOWARDS DESCRIBING OVERTONES IN ANIMATED NOTATION

Helen Svoboda
Monash University
helen.svoboda@monash.edu

Cat Hope
Monash University
cat.hope@monash.edu

ABSTRACT

This paper explores a set of notational possibilities for overtones on the double bass. Engaging animated notation approaches across two compositions, *Wormwood* (2022) and *Ice Bubbles* (2023), an imaginative solution is sought to address some of the issues and confusion surrounding traditional overtone scoring for string instruments. Photographs taken by the composer are engaged as sketches to form a starting point for both works, and as a source for visual information for the representation of particular overtone pitch sets and timbre. This is intended to create a more intuitive method for the depiction of instrument preparations and effects. In *Wormwood*, insect patterns photographed in the bark of a tree were drawn and put into motion on a computer. *Ice Bubbles* begins with a map of a floating ice island that is treated as territory for a boat journey from one shore to another. These starting points provide an overarching theme for the design of further details, resulting in two works that prompt new and unfamiliar creative decisions from the performer, responding to focussed compositional parameters in real time. This paper outlines the aims and processes of what we term the 'composer photograph' for a double bass performance of these works by the composer.

1. INTRODUCTION

Sketching has been used by humans to describe the visual world since prehistoric times [1]. As different tools have evolved, they have been adapted for sketching. Photography can be similar in "approach, process, speed of execution, and apparent honesty to nature to processes such as the 'plein air' sketch used in landscape painting" [2]. It is a process that makes the most of opportunity - what happens to be found or visible at a particular location can be captured and used as a basis for further development. In context of this paper, the term 'composer photograph' is used to describe this initial capture for musical purposes. The composer photograph is a photograph taken by a musician with the intention of using it as foundational material for a performance score. These photographs are used as frameworks for describing overtones in animated notation and a starting point for compositional explorations – seen in both *Wormwood* and *Ice Bubbles*. Once captured, their finer details are enhanced in Adobe Photoshop and put into motion to create an animated score.

Animated notation is a predominately graphic music notation that engages the dynamic characteristics of screen media [3]. It is particularly useful for representing sounds that are not served well by more traditional 'Common Practice Notations (CPN)' [4]. This includes music without a pulse, with electronic sound sources, or compositions where more performer input is desired. Further, screen media offers expanded possibilities for music notation, such as colour, dynamic motion, and the addition of other elements, such as audio tracks, into the score, creating a 'super score' [5].

1.1 Motion

The motion in animated notations can be generated and leveraged via different processes in these works. Preliminary eye tracking studies have demonstrated that musicians follow the motion in animated scores in a similar way to common practice notated works [6]. Bergrun Snæbjörnsdóttur's piece *2 Viti* (2012) [7] features a score that reveals itself gradually as the piece unfolds. The motion is followed by the performers, two guitarists in this case, and invites them to make decisions in real time. Known for creating "mutable, breathing, living structures through experimental performance practices and notation" [8], Snæbjörnsdóttur often projects scores onto the floor. This invites a sense of theatrical interplay visible to the audience. A focussed vocabulary of parameters is indicated in the score key, and the result is a slowly evolving, slippery sound world which would be impossible to notate any other way.

1.2 Pace of Change, Generative Form

Animating notations allows for the "revealing of sonic parameters in real time" [9] whereby the unfolding of a score can take on different elements of pacing as determined by the composer; either "randomly or planned" [9]. In animated notation, time may be tightly controlled across a set duration and specific time markers, but also lesser so in abstract forms. An example of the latter is seen in Reich's

Pendulum Music (1968); a text composition whereby "the piece's duration is dependent on the pace of the decelerating microphone trajectories" [10, p. 65]. It would be impossible to notate this motion in a way that is accurate, or in anyway useful, for each performance, however the piece provides an example of a work in which its pace and duration depends upon the velocity and speed of the objects to generate sound.

1.3 Timeline, Timeflow

Desainte-Catherine et al. (2013) define two conceptions of time in composition and performance – 'timeline' and 'timeflow', which are engaged in *Wormwood* and *Ice Bubbles*. 'Timeline' relates to chronology of musical events in the score itself, and the ways in which specific points of time correspondence or duration take shape in the 'authoring' of the music. The 'timeflow' represents data flow, relating to dynamic events that happen during the performance and the use of software to generate sequenced data in real time [11]. The interpretation of an animated score can exist across both paradigms, and as such, the combination of both can be incorporated into a work through the use of static data and dynamic events. The notational system in *Ice Bubbles* is split into two completely separated worlds: a static part filled with time-stamped static events, and a dynamic part filled with dynamic events handled by functions [12].

As such, there is potential for the combinatory fixed and active elements within animated notation to shift the performer's relationship with time, moving beyond numerical cues to incorporate focus on data flow. The use of both 'timeflow' and 'timeline' as corresponding paradigms can allow for a "unique musical experience with each performance, coupling the mercurial nature of improvisation with the more contemplative aspects of composition" [13]. An example is seen in Arne Eigenfeldt's *Unnatural Selection* (2014) – a work with challenging compositional parameters that generates new material across each iteration [14]. Subsequently, real-time interpretation of fixed score elements in generative animated notation such as this opens a variety of shifting possibilities across live performances.

As is discussed in the following sections, *Wormwood* and *Ice Bubbles* are works that operate within the time paradigms outlined above. In each work, generative pitch sets are interpreted in real-time. In *Ice Bubbles* the programmatic nature of the work brings multiple elements together to create an experience and reimagine a boat journey through a frozen lake.

1.4 Colour

Colour can be used in a wide range of ways in a musical score. Many composers use it to differentiate parts for the performer, or any variety of possibilities to the performer. Different shades denominate gradations in dynamics, timbre or other elements. An example of this 'musical literacy' [3] is seen in both *Wormwood* and *Ice Bubbles* whereby the use of colour not only represents pitch, but also timbre and density. This element can alter the way in which the performer responds to instruction more intuitively.

2. OVERTONE NOTATION

The prevalence of overtones in contemporary double bass repertoire exists across a limited and specialist area within composition and performance practices. The traditional and standardised notation system for overtones remains unresolved and spatially counterintuitive, leading to confusion amongst performance communities. In 'Harmonics on the Double Bass' (2010), Eric Daino argues:

> The twenty-first century presents many more options and variations for use [of overtones] by the performer and composer… In particular, there still remains a great deal of room for exploration and perfection of extended harmonics techniques on the double bass [15, p. 15].

A renewed interest in the overtone series post twentieth century has increased their prevalence in contemporary double bass repertoire across their natural and extended forms - yet this gradual evolution has somewhat furthered confusion around the use of overtones in regard to their performance and fingerboard placement, particularly across higher nodes; often the information becomes "too cluttered to adequately display what is necessary" [15, p. 4]. To counteract this ongoing issue, the two works discussed in this paper centralise the role of the performer/composer and subsequent creation of new work as a way of alternatively representing and communicating overtones in composition. They build on the work of leading performer/ composers in the field of overtone notation, including Fernando Grillo, Håkon Thelin, Stefano Scodanibbio and Mark Dresser – all of whom engage a variety of standardised overtone scoring methods in personalised combinatory ways [16]. Scodanibbio's *Sequenza XIVb* (2004) [17] uses a combination of tablature style notation, a double stave system and the 'o' symbol to denominate overtones in complex forms. Dresser's *K-Tude* (2010) uses partial numbers and the double-stave system to clarify pitch output across advanced harmonic nodes [18]. Both works highlight the variety of different ways in which complex overtone techniques can be combined to effectively communicate information to the performer. However, the limited range of extant methods commonly lack imagination in regards to the location of a harmonic node and its sounding pitch; especially as the techniques advance and become more finite and locationally specific.

Wormwood and *Ice Bubbles* seek to lessen this "discrepancy" [15, p. 2] by using graphic notation as a way to avoid the traps of standardised overtone notation, and move the focus on enabling a specific harmonic to look closer to how it sounds. Nodal numbers are derived from pitch sets across generative forms, focusing on the "physicality of music-making" [10, p. 66] as opposed to a set pitch output. Both compositions explore a visualisation of overtone relationships into an animated graphic format from an initial 'composer photograph', with the aim of creating a form of action notation that hones in on fingerboard placement and physicality, opening new creative pathways in improvisation, composition and performance.

2.1 Fingerboard Anatomy

As the use of overtone techniques in contemporary double bass repertoire continues to evolve, a thorough physical understanding of the fingerboard anatomy and problem-solving is commonly required prior to performance - particularly whereby the sounding pitch does not correlate with the fundamental. As discussed, traditional notation techniques have been innovated and combined in idiosyncratic ways in the work of leading performer/composers such as Thelin, Scodanibbio and Dresser; however the limitation of standardised overtone notation systems continues to pose issues in regards to the physicality of sound-making across this technical area.

2.2 Action Notation

Action-based notation can 'mediate' the relationship between body and instrument; akin to a 'choreography' of soundmaking [19]. In composition, these systems can be personalised in the way that they illustrate gestures upon an instrument where the focus is on spatial and body awareness, or "what to perform and how to perform it" [10, pp. 65–66] – strengthening a performer's relationship with their instrument and the operational aspect of realising sound. Australian composer Vincent Giles identifies a growing idiosyncrasy across modern notation systems post twentieth century, with an increased level of specificity across compositions and their individualistic ways of communicating information to the performer [19]. Aaron Cassidy's *Second String Quartet* (2013) is one such example where the score represents

> a physical mapping of the instruments in question in a type of tablature notation that shows the physical movement necessary in the piece – the performative gesture – rather than the resultant sonic output, which is variable [19, p. 1].

The indeterminate nature of the resultant sound is largely dependent on a list of variables including hand size and bow quality, with a greater focus on the notation movement itself as a part of the piece in its realisation. In a similar vein, Helmut Lachenmann's *Guero* (1988) [20] for solo piano incorporates a proportional topographic view of the instrument. Separated into three pitch registers, note heads of various shapes and fillings indicate the nature, direction and duration of the physical gestures. The resultant sounds are determined by the movement mapping embedded into the score (Figure 1):

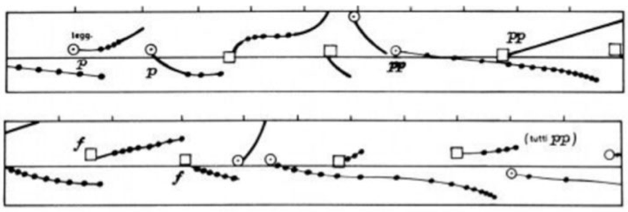


Figure 1. *Guero*. (1988). Physical gestures embedded in the score indicate sounds, techniques, and directions.

Similarly, *Wormwood* and *Ice Bubbles* explore the use of pitch maps and tablature style action-notation systems to enable a focus on what Kojs calls the 'physicality' of overtone production [10, pp. 65–66], prioritising eye-hand coordination and gesture at the centre of the sound. Instead of a focus on the desired pitch output, the performer is presented with information regarding the spatial position of a harmonic node; as such, the relationship with the fingerboard enters a visual, intuitive space in interpreting score materials. This creates what Kojs identifies as:

> music (that) is treated as a physical process, engaging our bodies and objects in actions. Experiencing music enactively - that is creating, notating, and performing it through such lens and ear - enriches our musical imagination and connects it to our everyday world [10, p. 71].

3. THE COMPOSER PHOTOGRAPH

The 'composer photograph' is used as a compositional framework in both *Wormwood* and *Ice Bubbles*. Upon the initial capture, the selection of visual material is a process that makes the most of opportunity; what happens to be found or visible at a particular location can be captured and used as a basis for further development within a score. If this capture is made/considered/found/identified by a musician, with the intention to sound the imagery, then this can be thought of as a composer photograph. Photographic details (i.e. colours, textures, shapes, lines) are accentuated and used as compositional foundations to intuitively represent overtones. These elements are used in combination with generative pitch sets which initiate a sense of overall pacing and flow to the performer.

Figure 2 outlines three phases of compositional development using the composer photograph as a method:

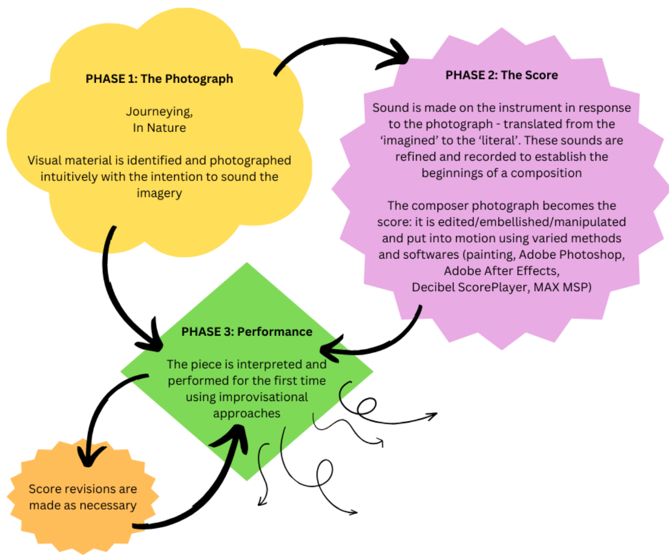


Figure 2. *The Composer Photograph*. (2024). Three phases of compositional development.

As shown in Figure 2, Phase 1 for these works takes place when journeying in nature. 'Journeying' here refers to time spent in new natural environments, where senses are heightened. This leads the composer to photograph visual material of interest with the intention of engaging with it as a score. Phase 2 involves the making of the score. Details are accentuated in Adobe Photoshop and put into motion (i.e. linear patterns, textures, colours) using software such as MAX [21] in the case of *Wormwood,* and the Decibel ScorePlayer [22] in *Ice Bubbles.* The performer ascribes sonic ideas that have formed in response to the visual stimulus by editing the image to be a score. In Phase 3, the work is interpreted for performance, responding to a combination of compositional approaches. The cycle begins again with refinements discovered as part of the performance process. As seen in Figure 2, the three phases create an iterative process, where the composers photograph provides a site for further score and performance development. This is described in more detail below.

4. WORMWOOD

Wormwood [23] is the first of a series of solo double bass works which tests the implementation of composer photographs of organic structures. Pitch sets are ascribed to a series of worm tracings in the bark of a tree, photographed in the Tasmanian forest in September 2022. The 'Glowing Edges' effect in Adobe Photoshop was applied to the images to enhance the linear patterns, zooming in on areas of interest (Figure 3):

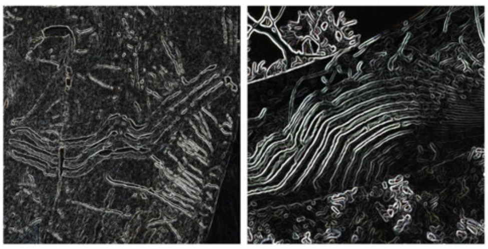


Figure 3. *Wormwood*. (2022). Worm patterns in tree bark, effected in Adobe Photoshop.

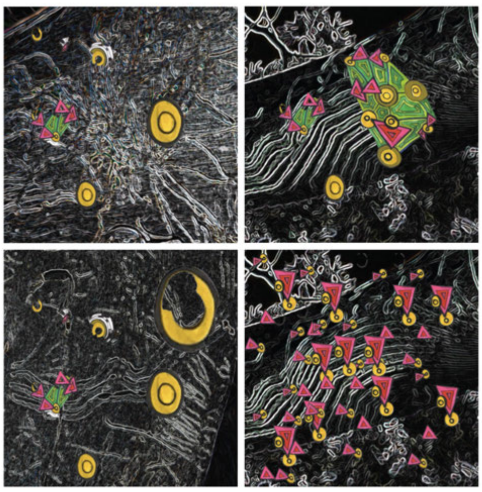


Figure 4. *Wormwood*. (2022). Preliminary experiments of overtone shape categories, drawn and overlaid over composer photographs of worm tracings in photoshop.

Multiple shape categories within the tree bark arose. These were drawn and overlaid upon the original image, then assigned different colour categories to indicate specific overtone relationships (Figure 4).

These images were then put into motion on a computer via a MAX patch. This generated changes and transitions in the shapes, creating an animated score. Variance in speed, motion, line and colour across each of the shapes indicates different arco techniques (ricochet, smooth, pulsating) and natural/artificial harmonics. Arco techniques are described through a left-to-right wobble of the shapes (ranging from slow to fast) and their smooth or jagged edges to represent articulation. Natural/artificial harmonics are represented in the multiplicity of the shapes; the first shape describes the fundamental, and each additional shape doubles the frequency and resulting pitch (Figure 5):

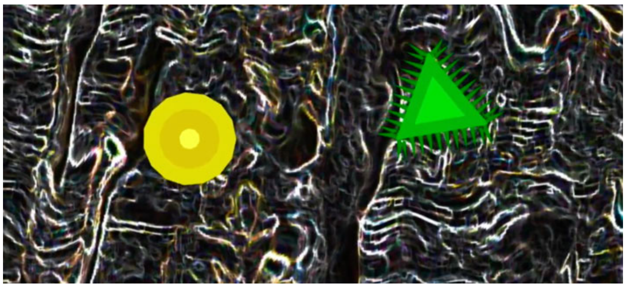


Figure 5. *Wormwood*. (2022). Screenshot of overtone shape categories in the animated score. The yellow shape indicates pulsating arco across a multiplicity of three (i.e. overtones 1, 2, 4). The green triangle indicates arco ricochet across multiplicity of two (i.e. overtones 3, 6).

Randomisation is implemented to break habitual improvisatory patterns with an increased sense of intention across the pitch sets. The use of animated notation subsequently alters the decision-making process into a more focussed state, opening a transitory space to discover alternate creative pathways in performance. The fluidity of the image evokes a sense of non-urgency, inviting more consideration across the parameters within the score key; colour represents pitch selection, and shape outlines and their corresponding movements represent articulation. The generative component of the score and these materials propels sound that is enlivened from the moving image - the sonic results of which rely on motion for its realisation [26].

5. ICE BUBBLES

Ice Bubbles [24] explores plucked overtones across an animated graphic score projected over the double bass, which is laying on its back. The piece illustrates a ten-minute boat journey from Suomenlinna island to Helsinki in the Finnish winter. The work is a sonic representation of the experience of daily ferry transit, during which the vessel would break through the frozen ice from shore to shore. Generative pitch sets and instrument preparations are subject to a variety of electronic effects that are described in the score. A composer photograph of an ice sheet forms the basis of the sketch and score materials.

The floating ice sheet is projected upon the double bass – or 'boat'. Four coloured gradients within the animated image represent shifting density, incorporating use of the

Chase Bliss MOOD MKII pedal [25] on the double bass to multiply sounds as the ice densifies using the delay and slip effects [26]. An identical gradient map appears in Decibel ScorePlayer software's canvas mode, in combination with a series of generative overtone pitch sets, to which the player responds. Instrument preparations incorporate aluminium foil, finger picks, scordatura and contact microphones to accentuate the tactility of crackling ice. Focussed contact microphone placement on the bridge further highlights the sonic evocation of crackling, capturing each individual overtone in both their raw forms and with use of effects to multiply these sounds and transform pitch as the ice thickens and cracks.

The score for *Ice Bubbles* comprises two parts, incorporating components of timeline and timeflow [11]:

1. Overhead Gradient Map/Floating Ice Sheet (projected onto the double bass) – *timeline*
2. Pitch Set Generator (read in the Decibel Score-Player software) – *timeflow*

5.1 Overhead Gradient Map / Floating Ice Sheet (timeline)

An overhead projection sees an animated image of frozen water projected onto the double bass, which is laid horizontally on its back – i.e. 'the boat' (Figure 6):

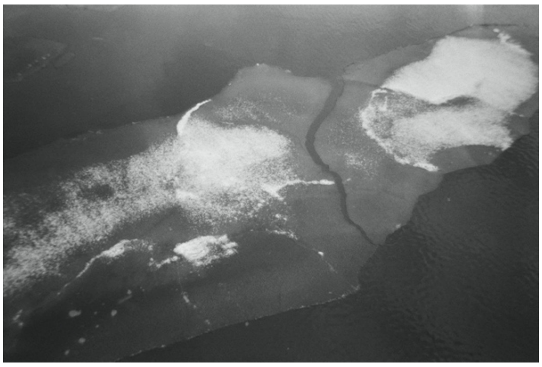


Figure 6. *Ice Bubbles*. (2023). Composer photograph.

Figure 7. *Ice Bubbles*. (2023). Pink gradient with aluminium foil. Indicates maximum density and use of FX.

Ice sheets within the photograph are highlighted with different coloured gradients (Figures 7, 8), to which the performer responds. Each of the coloured gradients indicates a varying degree of density and rate of change, resulting in a symmetrical compositional arc inspired by the nature of the frozen lake; closest to the shore, the ice is prone to melting due to more movement in the water, whilst the middle of the lake is thick and increasingly immovable.

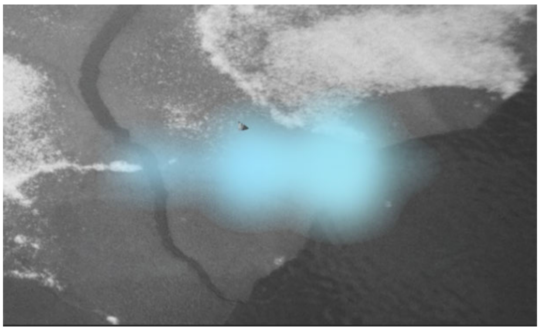


Figure 8. *Ice Bubbles*. (2023). Blue gradient with outline of the double bass (i.e. the boat). Indicates sparse density, without FX.

The gradient arc is explored across the following compositional parameters (see Table 1):

Gradient	Density
light blue → blue	Sparse → less sparse
Blue → purple	Becoming moderately dense
purple → pink	Moderately dense → Dense *** incorporates slip FX on MOOD MKII for multiplication + gradual addition of alfoil
deep pink → *alfoil*	Dense → Maximum density *** incorporates delay FX on MOOD MKII for multiplication + alfoil

Table 1. *Ice Bubbles.* (2023). Gradient parameters, representative of density, effects and preparations.

The amount of highlighted ice sheets increases towards the middle of the work as the ice densifies, then tapers off as the ferry moves back towards the shore. This element of the score is fixed, akin to the concept of a static timeline, but without a visible numerical duration. The pacing of the piece aims to evoke a sense of blurred time when travelling by sea.

5.2 Pitch Set Generator (timeflow)

Twelve pitch sets are generated in real-time within the Decibel ScorePlayer, using the canvas mode. The pitch sets appear across a series of randomised slides, to which the performer responds in combination with gradient information from the overhead projections. As the gradient begins to indicate more density, the pitch sets incorporate use of slip and delay effects on the MOOD MKII in combination with aluminium foil wrapped around the top two strings.

Pitch sets represent the four strings of the double bass, placed vertically in the score image frame (lowest to the left, highest to the right). Overtone numbers are placed within the coloured shapes to communicate desired pitch output and placement. Directional orange arrows indicate the ordering of the pitches (Figure 9):

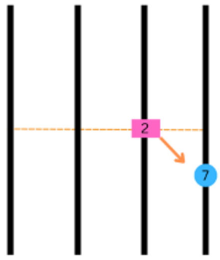


Figure 9. *Ice Bubbles*. (2023). Pitch Set #5 – the 2nd overtone on the 2nd string (in scordatura) is to be plucked with a fingerpick on the right hand index finger, followed by the 7th overtone on the G string with the left hand.

The generative form in this component of the piece incorporates the essence of timeflow to allow for dynamic real-time interaction with fixed materials [11]. The fixed gradient arc from shore to shore propels an abstract sense of time, conceptualising the journey without numerical cues. The use of effects allows for pitch transformations and multiplications across purple to pink gradients as the density grows, augmenting the metallic crackles from vibrating aluminium foil wrapped around the strings [33].

6. CONCLUSIONS

The use of the 'composer photograph' in both *Wormwood* and *Ice Bubbles* provides a framework for the creation of a new graphic notation system for overtones on the double bass. The visual information inherent within both initial sketches allows for a sense of zooming in; deriving information from the raw image to represent different parameters of sound, and ways in which to perform overtones. The incorporation of animation within both of these works creates a sense of dynamic and fixed data flow allowing for real-time interpretation of sequenced parameters. In *Ice Bubbles*, the overhead projection of the floating ice sheet remains fixed with the gradient arc, whilst the generative pitch sets are randomised to allow for interactivity and variance across performances. The notion of these two corresponding time paradigms within the programmatic nature of *Ice Bubbles* brings multiple elements together to create an experience and signify a boat journey through a frozen lake, resulting in an abstracted re-imagining of a distant memory. In *Wormwood*, the generative pitch sets appear on the screen, prompting creative decisions from the performer. The slowly morphing insect-like patterns on screen invite a sense of gradual motion and increased intentionality across each performance as it varies and evolves. The design principles across both works have informed alternative ways in which to communicate overtone techniques to the performer, inviting more intuitive ways to practice and guide the performance of overtones.

Acknowledgments

The Decibel ScorePlayer component of *Ice Bubbles* has been developed in collaboration with Aaron Wyatt. The video projection for *Ice Bubbles* has been developed in collaboration with Ciaran Frame. The MAX patch for *Wormwood* was developed by Ciaran Frame. *Wormwood* was premiered in Hamburg in 2022 as part of The Digital Score Project (DigiScore), which was supported by the Hamburg Institute of Advanced Study, the Hochschule für Musik und Theater Hamburg, and the Sir Zelman Cowen School of Music at Monash University.

7. REFERENCES

[1] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?" *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 1–10, 2012. https://doi.org/10.1145/2185520.2185540

[2] K. N. Ogden, "Musing on Medium: Photography, Painting, and the Plein Air Sketch," *Prospects*, vol. 18, p. 75, 1993.

[3] C. Hope, "Electronic Scores for Music: The Possibilities of Animated Notation," *Computer Music Journal*, vol. 41, no. 3, pp. 21–35, Sep. 2017. https://doi.org/10.1162/comj_a_00427

[4] R. Dannenberg, "Extending music notation through programming," *Contemporary Music Review*, vol. 13, no. 2, pp. 63–76, 1996.

[5] S. Emmerson, "Crossing cultural boundaries through technology," *Music, Electronic Media and Culture,* pp. 125–147, Routledge 2016. https://doi.org/10.4324/9781315596877

[6] L. Vickery and T. Goh, "Music Screen-Reading: indicative results from two pilot studies," in *ACMC2015 - MAKE!: Proceedings of the Annual Conference of the Australasian Computer Music Association*, pp. 119–125, 2015.

[7] B. Snæbjörndóttir, *2 viti*. 2012.

[8] B. Snæbjörnsdóttir. https://www.bergrun.com/about

[9] C. Hope, "The Future is Graphic: Animated notation for contemporary practice," *Organised Sound*, vol. 25, no. 2, p. 190, 2020. https://doi.org/10.1017/s1355771820000096

[10] J. Kojs, "Notating Action-Based Music," *Leonardo Music Journal*, vol. 21, pp. 65–72, 2011. https://doi.org/10.1162/lmj_a_00063

[11] M. Desainte-Catherine, A. Allombert, and G. Assayag, "Towards a Hybrid Temporal Paradigm for Musical Composition and Performance: The Case of Musical Interpretation," *Computer Music Journal*, vol. 37, no. 2, pp. 61–72, 2013. https://doi.org/10.1162/comj_a_00179

[12] M. Desainte-Catherine, A. Allombert, and G. Assayag, "Towards a Hybrid Temporal Paradigm for Musical Composition and Performance: The Case of

Musical Interpretation," *Computer Music Journal*, vol. 37, no. 2, p. 65, 2013. https://doi.org/10.1162/comj_a_00179

[13] A. Eigenfeldt, "Generative Music for Live Performance: Experiences with real-time notation," *Organised Sound,* vol. 19, no. 3, p. 276, 2014. https://doi.org/10.1017/s1355771814000260.

[14] A. Eigenfeldt, *An Unnatural Selection*. 2014. [Online] Available: https://aeigenfeldt.wordpress.com/an-unnatural-selection

[15] E. Daino, "Harmonics on the Double Bass," Honours Dissertation, University of Delaware, p. 15, 2010. [Online] Available: https://udspace.udel.edu/handle/19716/5505

[16] H. Thelin, "A Folk Music for the Double Bass," 2010. [Online]. Available: https://haakonthelin.com/multiphonics/a-folk-music-for-the-double-bass

[17] L. Berio, S. Scodanibbio, *Sequenza XIb. 2004*. Universal Edition.

[18] M. Dresser, *Guts: Double Bass Explorations, Investigations, Explanations*, Kadima Collective Recordings, 2010.

[19] V. Giles, "Scoring For Grid Controllers – Preliminary Considerations Of Action, Gesture, Functionality, And Limitations" in *Australasian Computer Music Conference*, Sydney: Australasian Computer Music Association, 2015.

[20] H. Lachenmann, *Guero: für Klavier = for piano*. Wiesbaden: Breitkopf & Härtel, 1988.

[21] "MAX", Cycling 74. [Online]. Available: https://cycling74.com/products/max

[22] "Decibel Score Player", Decibel New Music Ensemble. [Online]. Available: https://decibelnewmusic.com/decibel-scoreplayer

[23] H. Svoboda, Performer and Composer, *Wormwood* [Video]. Hamburg Institute of Advanced Studies, 2022. Available: https://youtu.be/ANoWI23Z7Rw

[24] H. Svoboda, *Ice Bubbles*. 2023.

[25] "MOOD MKII," Chase Bliss. https://www.chasebliss.com/mood-mkii

[26] H. Svoboda, Performer and Composer, *Ice Bubbles* [Sound Recording]. Melbourne, Australia, 2023. https://on.soundcloud.com/bYXGJ

FORMS: GENERATIVE VISUAL MUSIC THROUGH GRAPHIC NOTATION AND SPECTROGRAM SYNTHESIS

Santiago Vilanova Ángeles
Independent Researcher
santi@playmodes.com

ABSTRACT

This paper presents an in-depth analysis of FORMS, an extensive generative visual music research project that blends graphic notation with spectrogram synthesis. Rooted in the principles of generative music, FORMS encompasses a broad spectrum of digital art projects, unified by a unique software framework. The research bifurcates into two pivotal domains: spectrogram synthesis via digital image sonification and graphic notation tailored for musical instruments. This comprehensive study sheds light on the transformative capabilities of visual-to-auditory conversions in contemporary generative music, providing novel perspectives on animated algorithmic notation systems and some of their possible artistic applications.

1. INTRODUCTION

Embarking on a journey through generative [1] graphic notation and spectrogram synthesis, FORMS fuses digital art with contemporary music practices. Incepted amidst the 2020 pandemic lockdown, this initiative has since birthed a number of eclectic projects, ranging from online streaming bots to audiovisual installations or string quartet performances. Central to these projects is a bespoke software system, powering the expressive and aesthetic potential of algorithmic Visual Music [2] notation systems.

This article narrows its focus on the intricacies of the FORMS software systems and their diverse artistic manifestations. Broadly, the tools and their resultant projects can be categorized into:

- Spectrogram synthesis and image sonification. [3]
- Graphic notation tailored for instruments.

A unifying theme across these categories is generativity, with each FORMS project incorporating elements of randomness and probability to achieve unpredictable musical compositions.

This article also presents several case studies, comprising visual and musical compositions by the author, to illustrate the discussed concepts and techniques.

The research showcased in this article spans four years and encompasses the development of mathematical techniques, software algorithms and tools, audiovisual installations and instruments or contemporary music concerts. Given the breadth and depth of this work, each aspect could warrant its own detailed article. In fact, the author is currently working on a PhD thesis to comprehensively address all facets of this research. However, this article aims to provide a concise overview of the project and its recent artistic achievements.

Many of the software developments that are described in the article have been granted open access through online repositories, which will be provided punctually.

Along the text, you will also find a number of footnotes which point to website links showcasing video documentation of the described elements or artworks. We encourage the consultation of these links for a proper understanding of the research presented here, given its animated nature.

2. SPECTROGRAM SYNTHESIS AND IMAGE SONIFICATION

While traditionally a spectrogram serves as a representation of pre-existing sound, within the realm of FORMS it is the genesis of music. This paradigm shift involves first crafting the visual representation of sound, followed by the application of image sonification and sound synthesis techniques.

In order to effectively create spectrograms that 'sound good' once sonified, we first defined a number of rules for the image generation and its subsequent sonification, which establish the basic axioms to take into account:

- The X-axis denotes time, *sonogram units* positioned along it to express musical rhythms. The resulting images can either scroll horizontally towards a static play header or be traversed by the header itself.
- The Y-axis represents pitch, where each pixel row aligns with a microtone within the overall pitch range, or *tessitura*. A higher position of a pixel row indicates a higher corresponding pitch. Given that our measurement is based on pitch rather than frequency, the generated images conform to the principles of Mel-spectrograms [4].
- Each pixel row is associated with a sinusoidal sound wave. The frequency of this wave is determined by the tessitura and the total number of microtones,

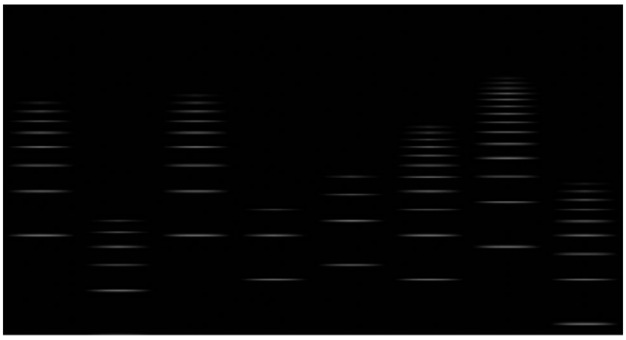


Figure 1. An example of partial clustering, to design bass sounds.

which is equivalent to the height of the sonified image in pixels. For instance, if an image stands at 120 pixels in height and covers a tessitura of one octave (12 semitones), this results in 10 microtones per semitone, calculated as 120/12.

- Luminance determines volume, with darker shades indicating lower volumes and brighter shades denoting higher volumes.

Considering these guidelines, we developed a comprehensive collection of procedural sonogram units, as we will see in section 2.2.2. Drawing upon Fourier Transform principles, we can synthesize intricate harmonic content by layering partials and adjusting their amplitudes through assigned luminance variations (Fig. 1). This approach enables the precise visual representation and synthesis of any sound by detailing its partials on the Mel-spectrogram.

This image sonification technique can be perceived as a variant of additive or spectral synthesis [5], where an array of sinusoidal waves, equivalent to the image's pixel height, plays concurrently guided by a play header that traverses the image. The luminance data from the analyzed pixel column serves as the amplitude bins for the sinusoidal wave bank, re-scaling the 0–255 luminance range to a sound linear amplitude range between 0 and 1. In turn, the frequency bins are ascertained by each pixel row's index after performing a pitch-to-frequency conversion. Please see section 2.2.2 for an in-depth review of FORMS image sonification methods.

In regards to time sequencing, and using the grid depicted in Fig. 2 as a foundational layout, we place the aforementioned sonogram units on the canvas. By incorporating inherent randomness into generative music, the system determines the presence of specific sonogram units (e.g., a kick drum, a bass note, a bell tone) at particular positions on the time axis, thanks to a series of generative sequencing algorithms that have been also developed and integrated into the software system. This same generative approach has been implemented for the pitch height of sound objects, giving the software system a certain range of freedom after the selection of a musical scale. Please see section 2.1 for details on the implemented generative music techniques.

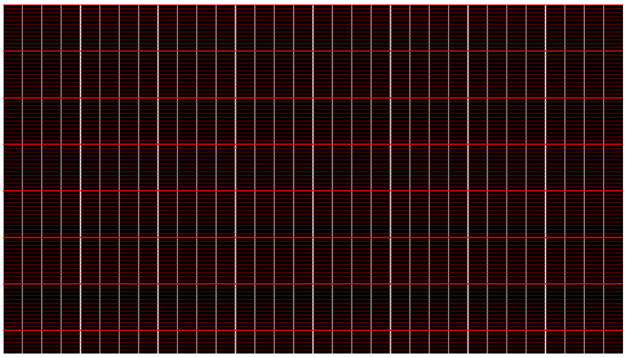


Figure 2. The sonification grid. Horizontal red lines denote steps on the chromatic scales, the thicker ones being the note C of successive octaves. White vertical lines denote quarter notes, the thicker ones being the start of a bar.

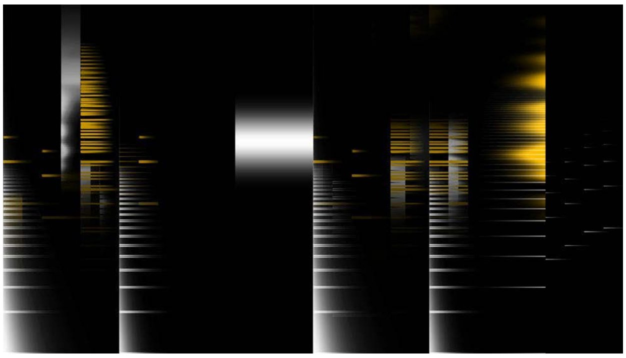


Figure 3. Generatively synthesized spectrogram showcasing varied percussive elements. Observe the multi-partial bass aligned with the kickdrum and the distinct melodic arpeggio in the final bar.

All in all, this approach crafts a realm of potential resulting images [1] , as seen in Fig. 3.

2.1 Generative Music Algorithms

At the heart of the spectrogram synthesis system lies a number of principles related to musical generativity. This is understood as the capacity for automatic music generation through rule-based processes, controlled randomness and other algorithm-based automated procedures. Our objective has consistently been to ensure that every software system within FORMS incorporates algorithms that enable it to operate as a musical automaton. By interpreting user-defined rules, the system can produce endless variations within the confines of these guidelines. To achieve this, we've explored and integrated various generative music techniques, laying the foundation for this automated music generator. Below, we delve into the most pivotal of these techniques:

2.1.1 Euclidean Rhythms

Euclidean rhythms [6] are a unique approach to rhythm generation based on the Euclidean algorithm, which deter-

[1] the following video playlist displays a collection of FORMS spectrogram synthesis results: https://shorturl.at/JOQZ2

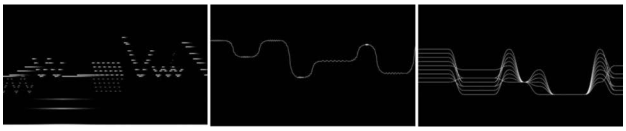


Figure 4. Examples of harmonic sequencing, featuring arpeggiation, root progressions and chords.

mines the greatest common divisor of two numbers. By distributing a set number of beats as evenly as possible over a rhythmic cycle, intriguing and musically relevant patterns emerge. These rhythms have been identified in various traditional music cultures worldwide, making them a versatile tool for generating diverse rhythmic structures in contemporary compositions.

2.1.2 Probabilistic Sequencing

Probabilistic sequencing introduces an element of chance into the sequencing process. Instead of following a fixed pattern, each step in the sequence has a probability associated with it, determining its likelihood of being played. This approach allows for dynamic and ever-changing musical patterns, ensuring that compositions remain fresh and unpredictable, yet within defined boundaries set by the composer.

2.1.3 Generative Harmony

Various generative harmony techniques have been incorporated. Users have the flexibility to select a musical scale for each synthesized spectrogram from a diverse list, which spans from ancient Greek modes to contemporary scales and world-music modes. Based on this selection, random chord progressions are generated by stacking notes at random intervals. While there's room for further refinement, such as the incorporation of functional harmony, serial and tone-row tonal systems – just to name a few –, the modal harmonies produced by this approach give a first taste of pitch organization within the FORMS system. Additionally, a user-defined weighted chance algorithm determines the likelihood of the composition shifting its root note to any given degree or undergoing a specific semitone transposition. This means that each scale degree and every chromatic step have individually assignable probabilities. The culmination of these features produces spectrograms with dynamic and scale-quantized harmonic progressions (Fig. 4).

2.1.4 Mute/Solo Chance

Considering that compositions frequently consist of multiple layers (such as percussive elements, melodies, chords, arpeggios, bass or textures), an automated mute/solo feature has been integrated. This functionality either emphasizes specific layers (solo) or suppresses them (mute), adding rhythmic depth to multi-layered compositions. Importantly, the mute/solo algorithms are cognizant of time positioning, allowing the probability of a mute/solo event to be influenced by the composition's temporal context. For instance, a solo might be more likely during the final bar of an eight-bar sequence.

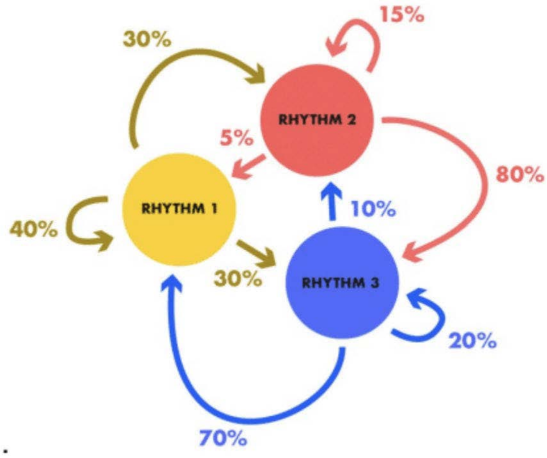


Figure 5. An example of 1st order Markov Chains application for rhythmic transitions.

2.1.5 Algorithmic Arpeggios

Algorithmic arpeggios utilize mathematical formulas to craft arpeggiated sequences, deconstructing chords into a series of notes played consecutively. The system boasts a parametric approach to arpeggiation, encompassing variables like pattern length, direction, polyphony, tempo, and interval randomness. By assigning random values to these parameters, the system can generate a limitless array of dynamic and evolving arpeggios (Fig. 4).

2.1.6 First Order Markov Chains

Markov Chains serve as a potent instrument for intelligent sequencing. Primarily used for navigating between presets (where a preset refers to a specific set of parameters defining a spectrogram), Markov Chains facilitate the determination of transition probabilities from one state to another in a structured manner (Fig. 5). Within a musical framework, this mechanism proves invaluable for guiding chord transitions, tonal shifts, and stylistic evolution.

2.1.7 Noise Functions

Noise functions, commonly utilized in computer graphics, have found applications in musical domains as well. By harnessing algorithms such as Perlin noise [7], we generate a sequence of values that produce organic and evolving control signals. These signals can be channeled to modulate various parameters, such as the volume of individual instrumental layers, leading to a natural interplay of voices. They can also influence the rate of LFOs that adjust volume or the number of sound partials, or even the tempo of an arpeggio. Consequently, compositions acquire a sense of organic fluidity and unpredictability, enhancing their overall appeal.

2.2 Software Implementation

The applications underpinning spectrogram synthesis and sonification are developed across multiple coding platforms, interconnected through the OSC network protocol.

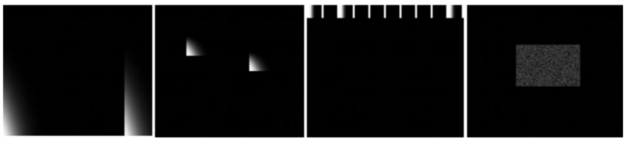


Figure 6. Examples of percussive sonogram units. Kickdrum, Snaredrum, Hihat, Noise.

Although the software architecture remains in a prototypical phase, with aspirations towards a unified application handling all processes, its current iteration has demonstrated robust functionality. This makes it suitable for practical applications, such as audiovisual installations and live concerts. The entire software infrastructure, as illustrated in Fig. 8 and Fig. 7, is optimized for real-time execution, establishing it as a tool for live generative visual music performances.

2.2.1 Image Synthesis with Processing

Image synthesis is executed using Processing [8], an open-source Java-based framework tailored for procedural graphic design. We developed multiple functions that define diverse bitmap graphics generation algorithms. As we explained in section 2.1, all these functions integrate a degree of controlled randomness, ensuring the outcomes are never identical, aiming for a generative music system.

At the center of the spectrogram synthesis methodology lies the concept of the *sonogram unit*. Various functions have been developed to define these visual entities. Each function can generate a segment of a spectrographic image, which can then be strategically positioned on the canvas, aligned to a time or pitch grid, or overlaid with other sonogram units through additive or alpha visual blending techniques. A plethora of these sonogram unit functions have been crafted, encompassing elements such as kick drums, snare drums, hi-hats, bass, bells, and pads. Typically, these functions can be customized with parameters like x and y coordinates, duration, or specific attributes tailored to each object, such as the number of partials for elements that adhere to the harmonic series. Leveraging the generative sequencing algorithms outlined in section 2.1, these sonogram units are rhythmically and tonally arranged on the canvas, leading to a boundless array of musical expressions.

Typically, the final output from this Processing application is a FullHD image, 1920x1080 pixels, saved as a .png file. The time required for this software system to synthesize a new image upon request varies between 1/2 to 5 seconds, contingent on the graphic algorithms' complexity and the available CPU.

2.2.2 Image Sonification with Max/MSP

Image sonification is managed by a custom Max/MSP [9] patch. This patch retrieves the path of the newly generated image from Processing, loads the image into memory, and performs a luminance analysis on a specific column of pixels, referred to as the *header*. The header can either move from the image's left to right at a set BPM or remain fixed while the image scrolls from right to left. In both scenarios, the movement of the header or the image is driven by an audio rate phasor, quantized to pixel column positions.

The results of the header's luminance analysis are used to adjust the amplitudes of a predetermined frequency bank of sinusoidal oscillators. The frequency values of these oscillators are based on the user-defined tessitura, while the bank's size (and the number of microtones within the tessitura) corresponds to the image's pixel height. As the header progresses across the image, the amplitudes of the oscillator bank evolve, producing a sonic representation that mirrors the visual content of the image. Following a simple audio mastering process that includes limiting and subtle reverb adjustments, the resulting sound is channeled to the computer's audio output. From there, it can be further manipulated on external mixers or directly relayed to speaker systems.

2.2.3 Control Architecture in Max/MSP

Also acting as a conductor, the Max/MSP patch requests Processing for new images, transmitting an OSC data bundle with details about the graphic methods to be employed, the image's bar count, the musical scale to be utilized, among other characteristics. The user-friendly GUI design enabled by Max/MSP has facilitated the creation of an interface where users can precisely define the details of the spectrograms to be synthesized. A preset system simplifies the storage of this features. Each preset will consequently produce an infinite collection of images (owing to the integrated randomness in the Processing graphic algorithms), which, once sonified, will give rise to varied musical streams, from raging rhythms to pleasant harmonic ambient or complex melodic sequences.

2.2.4 Image Scroller Coded in C++

The final software component in this suite is the *Scroller*. Developed in C++, this straightforward application manages the scrolling of the .png format spectrogram images generated by Processing, aligning with the speed and phase set by the audio rate phasors in Max/MSP. Configurable for full-screen display, the Scroller is particularly advantageous for public concerts or installations due to its efficient GPU-based operation.

2.2.5 Open Source Software Repository

All the developments described previously on this section have been made public on a Github repository [2] .

2.3 Case Study: FORMS – Screen Ensemble

Forms – Screen Ensemble [3] (Fig. 9) made its debut at the 2020 edition of the Ars Electronica Garden [10] in Barcelona, courtesy of a grant from New Art Foundation [11], Institut Ramón Llull [12], and Hangar.org [13]. It is conceptualized as a generative visual music jukebox triptych. Governed by chance and probability, this automaton fabricates endless, unique spectrograms which are instantaneously converted into sound via sonification algorithms,

[2] https://github.com/PlaymodesStudio/FORMS
[3] https://www.playmodes.com/home/forms-screen-ensemble

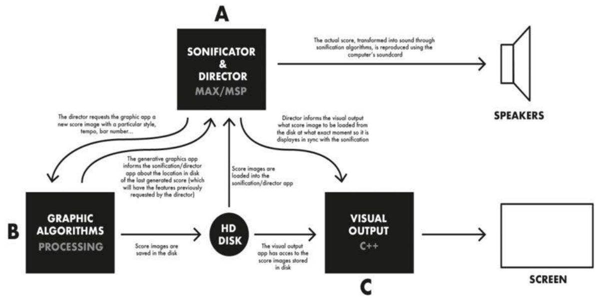


Figure 7. A diagram showing the data flow between the different software applications that conform the spectrogram synthesis system.

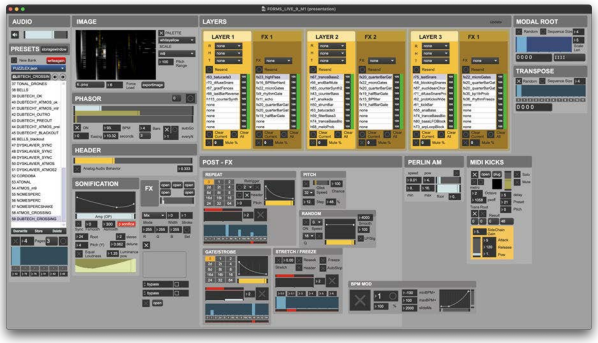


Figure 8. The FORMS Max/MSP patch, showing the GUI with the parameters that control the spectrogram synthesis.

enabling listeners to audibly perceive what they visually observe. Each screen within this interconnected ensemble assumes a specific instrumental role, be it Rhythm, Harmony, or Texture. Inside each of these 3 screen boxes, there is a FullHD monitor, a soundbar speaker, a computer and an Ethernet switch which interconnects the 3 computers into a single network (Fig. 10). Orchestrated by this trio of automatons, a visual music symphony continually evolves, giving rise to singular sonic landscapes that will never be replicated: from tonal ambient music to intense rhythms, surreal electronic transitions, or dance-floor beats.

Extensive mathematical research has been undertaken to code timbrically and musically coherent graphics that generate captivating rhythms, harmonies, or textures. Additionally, a suite of sequence control routines governs the compositions' evolution over time, leveraging weighted probabilities, Markov chains, and chance algorithms. Its generative nature yields endless, unique, non-repetitive, and rich music, liberated from the confines of conventional written musical language.

2.4 Case Study: Sonògraf

The *Sonògraf* [4] is conceptualized as an electronic audiovisual instrument. Designed as a musical learning tool for primary schools, it facilitates the transformation of handmade drawings or object collages into music, converting gestural strokes and geometric figures into electronic sounds. A collection of buttons and potentiometers enables live manipulation of the drawing's sonification characteristics, allowing users to accelerate, decelerate, or pause the resultant music and also choose its scales and tonalities. The Sonògraf is equipped with speakers and a compact video projector, enabling both intimate practice and staging in an audiovisual concert for a diverse audience. The conception, engineering, and design of the Sonògraf have been meticulously crafted to ensure intuitive play for all, offering a novel paradigm for understanding musical notation and deviating from the academic nature of traditional notation based on staves.

The Sonògraf was awarded the "People's Choice Award" and a jury honorable mention at the 2024 edition of the Guthman Musical Instrument Competition [5] held by Georgia Tech University, in Atlanta. A performance with the instrument, and in collaboration with clarinet player Lauren McCall, can be seen in a recording published by Georgia Tech. [6]

Figure 9. A photo of the Screen Ensemble as exhibited in Arts Santa Mònica during 2020's Ars Electronica Garden.

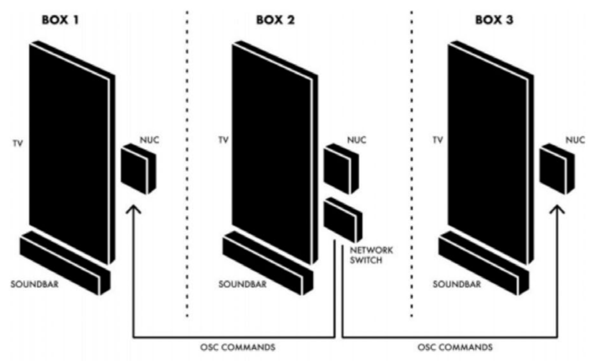


Figure 10. A diagram showing the different devices that compose the Screen Ensemble piece, and its network connections.

[4] https://www.playmodes.com/home/sonograf

[5] https://guthman.gatech.edu

[6] https://www.youtube.com/watch?v=j4gHsl5CQRs

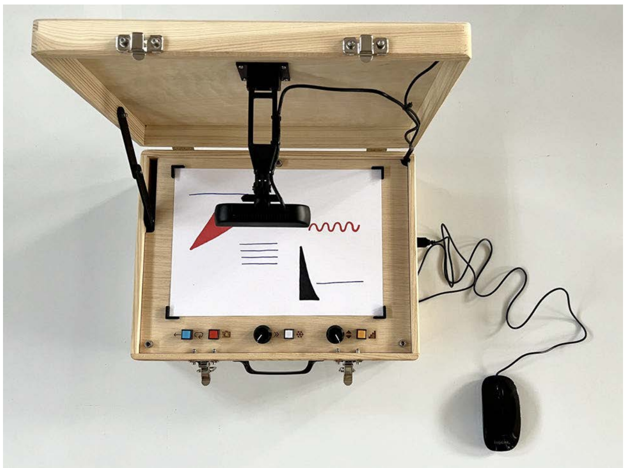


Figure 11. The “Sonògraf”, an educational tool for primary schools oriented towards music learning through hand-drawn, automatically sonified, graphic scores.

Finally, a Sonògraf DIY guideline has also been made public – using a Creative Commons license – on a Github repository. [7]

3. GENERATIVE GRAPHIC NOTATION FOR INSTRUMENTAL INTERPRETERS

Diverging from the spectrogram synthesis techniques, the FORMS initiative also delves into generative graphic notation. The primary distinction between these methodologies is that while spectrogram synthesis is predominantly employed to craft digital electronic music, the generative graphic notation techniques facilitate the creation of simple and succinct notation for human instrumentalists.

The generative notation approach within the FORMS series aims for an integration with the results of the spectrogram synthesis. The aim is to amalgamate both methodologies into multi-layered musical compositions, blending live instrumental music with sonified synthetic spectrograms. In anticipation of this integration, the rules governing the generation of graphic notation closely mirror those of spectrogram synthesis: Time is represented on the X-axis of the image, while Pitch is encoded on the Y-axis. However, a fundamental deviation from the spectrogram synthesis technique is that the volume or dynamics of the composition are not represented using luminance but through shape weights12. Additionally, to ensure clarity for the interpreters, when tonal precision is mandated in the composition, an annotated text preceding each musical object indicates the exact pitch to be played. A color code is also employed to distinguish different instruments, allowing each instrumentalist to identify their part without necessitating individual scores.

3.1 Software Implementation

The creation of the graphic notation scores is executed exclusively using Processing (Fig. 13). As the image sonification is performed by human interpreters, a software algo-

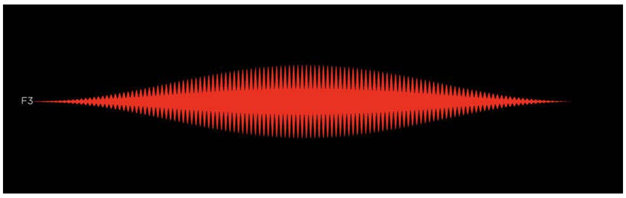


Figure 12. Dynamic volume contours are defined through the weight of the shapes. Note the accurate dynamic progression from ppp>ff>ppp and the fast tremolo contour. For tonal precision, the English notation is added at the beginning of the shape.

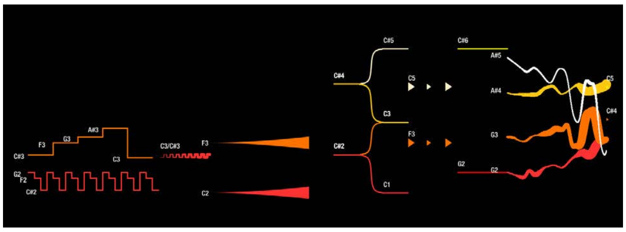


Figure 13. A screen capture of the Processing generative scoring software for string quartet.

rithm for image-to-sound synthesis is redundant. The approach for this set of graphic algorithms diverges from the previous spectrogram synthesis technique in that the generated graphics are vector-based, as opposed to bitmaps. Concurrently, each musical piece to be composed is treated as a unique project. While randomness and probability remain integral, each musical passage is meticulously programmed and its final visual output is curated from the multitude of musical/visual outcomes produced by the integrated randomness factors. The final selected .svg files are subsequently edited using a vector graphics editing software (e.g., Adobe Illustrator or Inkscape), ensuring macro-structural coherence for the entire composition. These files are then animated using an animation or post-production suite (e.g., Adobe After Effects).

Currently, efforts are underway to develop a new software system for macro-structured real-time generation of graphic notation, tailored for adept human interpreters. This new set of software tools would free us from the need of manual curation of results and its subsequent manual edition and animation. It is anticipated that in the upcoming months, a series of concerts will be made possible with these novel tools.

3.2 Case Study: FORMS – String Quartet

A collaboration with Quartet Brossa [14], was showcased at Barcelona’s CosmoCaixa Science Museum [15] in April 2021, as part of the art+science NEO cycle [16], curated by Irma Vilà. The project received support from the Department of Culture of the Generalitat de Catalunya [17].

FORMS – String Quartet [8] (Fig. 16) is a live multimedia concert designed for a string quartet, electronic music, and panoramic visuals, rooted in the concept and techniques of

[7] https://github.com/PlaymodesStudio/Sonograf

[8] https://www.playmodes.com/home/forms-string-quartet

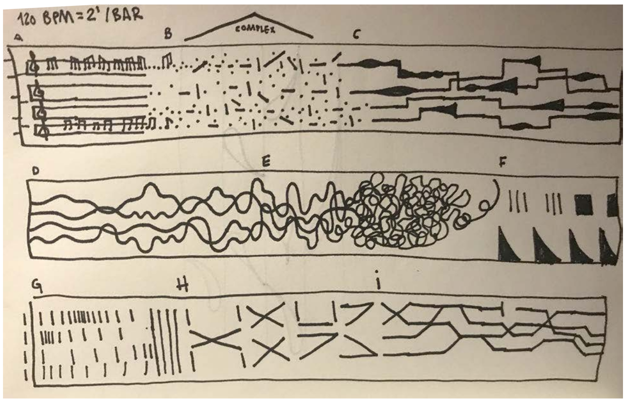


Figure 14. A preparatory hand-drawn sketch for the FORMS – String Quartet concert.

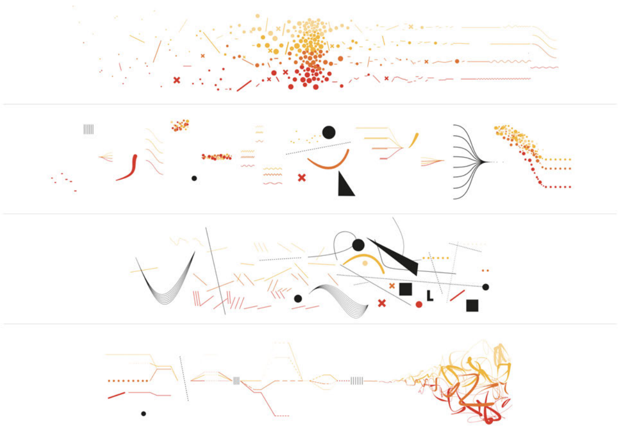


Figure 15. The final score of the FORMS – String Quartet's first movement. Each of the 4 warm colors is the score for the quartet: cream, first violin; yellow, second violin; orange, viola; red, cello. Black is electronic music, to be sonified using the methods discussed at 2.2.

image sonification. Musicians interpret a series of graphic scores, which concurrently construct the visual scenery, offering the audience an immersive experience where they can anticipate the 'music to come'. These scores, in conjunction with the generative spectrograms that form the electronic accompaniment of the piece, have been crafted through the use of the software tools seen in section 2.2. The composition process for this concert diverged from traditional methods. Following the initial commission, a decision was made to assign distinct colors to each instrument's score part so that all four melodic lines of the string quartet could coexist on a single drawing. This idea was first visualized through hand-drawn preparatory sketches, as shown in Fig. 14. These sketches were subsequently modeled algorithmically in Processing, as depicted in Fig. 13, with added randomness to generate varied outcomes for each section. The scores, produced by Processing in .svg format, were curated and served as the foundation for the macro-structural composition, developed using standard vector graphic editing software. The final vector score (Fig. 15) was animated using video editing software, ans resulting in a panoramic format video. Two versions

Figure 16. A photography of the 2021 *FORMS – String Quartet* premiere concert at the "NEO" art and science cycle in Cosmocaixa, Barcelona's science museum.

of this animated score were produced: a high-resolution panoramic version for concert visuals, and a 16:9 full HD version for musicians' reference. It's noteworthy that the musicians' version employed a distinct color scheme for clarity. While the scenic video displayed instruments in four shades of red, the musicians' version utilized a varied palette: red for the cello, blue for the viola, green for the second violin, and yellow for the first violin.

Alongside the string quartet score, a layer of visual electronic music was synthesized. This synthesis was based on the methods of spectrogram synthesis detailed in section 2 and was synchronized with the string quartet components.

Instead of traditional paper scores on a lectern, musicians refer to a 45" TV screen during rehearsals and concerts. This screen displays the animated score in synchronization with the panoramic scenic version.

The concert, which has a duration of 25 minutes, explores the expressive power of generative graphic notation, and deploys an extensive research on instrumental techniques – and its representation – from the classical legato or pizzicato, to extended techniques such as col·legnos, glissandi or Bartok pizzicati.

Since its premiere in 2021, this concert has been represented dozens of times in auditoriums and contemporary music festivals, and it has been widely acclaimed by audiences and specialized press.

4. "MIDIFICATION" OF GENERATIVE GRAPHICS

Within the FORMS framework, another avenue of exploration is the conversion of images to MIDI. The scoring rules mirror those of the spectrogram synthesis and graphic notation techniques (X-axis represents time, Y-axis denotes pitch, with luminance serving as the metric for volume dynamics. Review sections 1 and 3). Given the event-based nature of the older MIDI protocol [18], where MIDI velocity remains static during the playback of a note, luminance gradients cannot be employed to continuously modify volume expressions. Thus, dynamics are fixed at the onset of each note. Although newer iterations of the MIDI

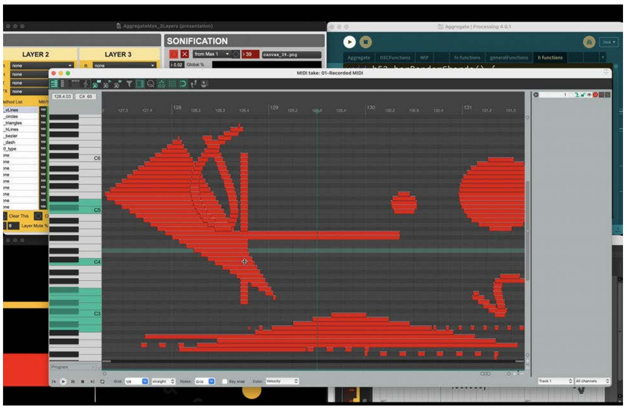


Figure 17. A screen capture of the resulting MIDI data from the sonification of generative graphics.

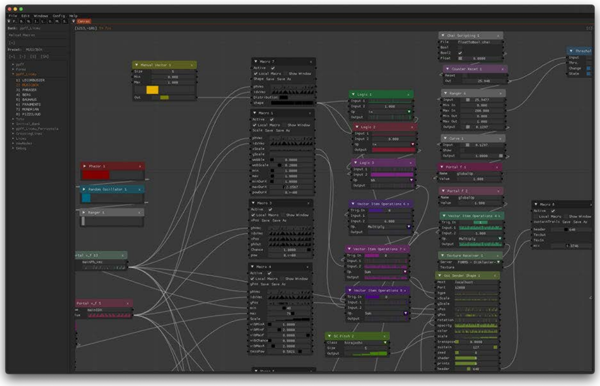


Figure 18. *Oceanode,* the nodal development environment software developed by Playmodes, which controls the graphic score generation of the *ppff* installation.

protocol (MIDI 2.0 and MPE) permit such continuous dynamic control along the duration of a note, the majority of existing MIDI instruments have yet to adopt these features.

4.1 Software Implementation

The MIDI-centric generative visual scoring tools are the latest additions to the FORMS suite. This set comprises two applications. Firstly, a nodal programming environment developed in Open Frameworks/C++ [19], responsible for defining shape types (e.g., circles, triangles, rectangles) and their XY positions, sizes, luminance, etc., for a maximum of 4096 shapes (Fig. 18). This nodal environment, called Oceanode, can be found as an Open-source tool at Playmodes' GitHub repository [20]. Secondly, a custom Processing/Java application processes the OSC data bundles from the C++ nodal environment, rendering this data as a visual output in real-time. Processing also oversees the analysis of luminance crossing the read header, converting this data into MIDI events.

This novel, streamlined approach has proven to be highly efficient and adaptable, setting a new benchmark for the future of the FORMS graphic generation software.

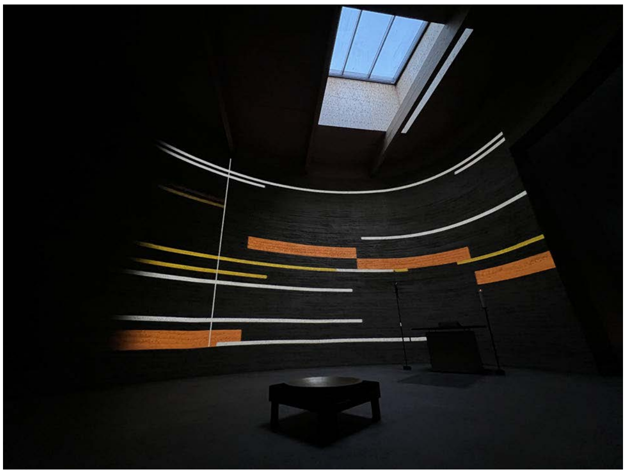


Figure 19. An image of the *Clash/Blend* installation at the Reconciliation Chapel in Berlin, 2022. The scores scroll towards the organ position and are sonified once the graphics cross the vertical header.

4.2 Case Study: Clash/Blend

Clash/Blend [9] (Fig. 19) is a visual music installation that commandeers the pipe organ at the *Kapelle der Versöhnung* in Berlin. Featured in the 2022 edition of the *Aggregate – New Works for Automated Pipe Organs* [21] festival program, this installation aligns with the Color Organ tradition. It deploys a video-projected visual canvas on the chapel walls. A cascade of geometries scrolls towards the Chapel's pipe organ, undergoing a transformation into MIDI (see Fig. 17) as these drawings intersect with the organ's position. Each color on the score is routed to a different MIDI channel, activating different sound registers per color. Owing to its generative nature, these graphic scores, and the sounds they produce, are unique and non-repetitive. This installation was developed using Processing (for real-time image generation) and Max/MSP (for MIDI translation of images), as taught in section 4.

in 2023 *Clash/Blend* was awarded a Golden LAUS [22] design prize, the most prestigious recognition in Spanish design.

4.3 Case Study: FORMS – ppff

Premiered on July 2023 at Barcelona's *Liceu* Opera [23], *ppff* [10] – an acronym for the musical notation *pianissimo-fortissimo* – is a generative visual music installation. The centerpiece of this installation is a *Steinway and Sons Spirio* player piano, a MIDI-automated grand piano capable of striking its hammers in response to real-time music data. This piano derives its musical input from an unconventional source: a continuous cascade of generative graphics (Fig. 20), showcased on a vertical LED screen. As these graphics reach the piano's position, they are transmuted into MIDI notes, facilitated by Oceanode, the aforementioned set of custom software developed in C++ (Open Frameworks), and Java (Processing) (See section 4).

[9] https://www.playmodes.com/home/clash-blend
[10] https://www.playmodes.com/home/ppff

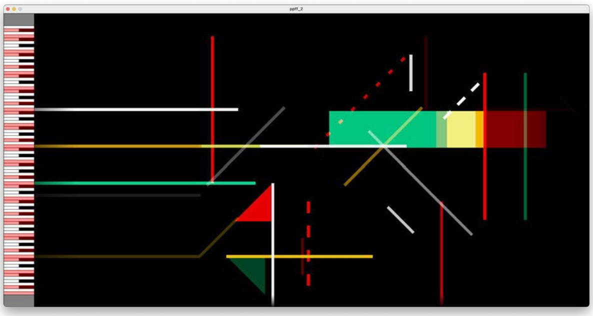


Figure 20. an example of the resulting graphic scores of ppff. Note the keyboard diagram at the left, which shows the positioning of the image-to-midi conversion.

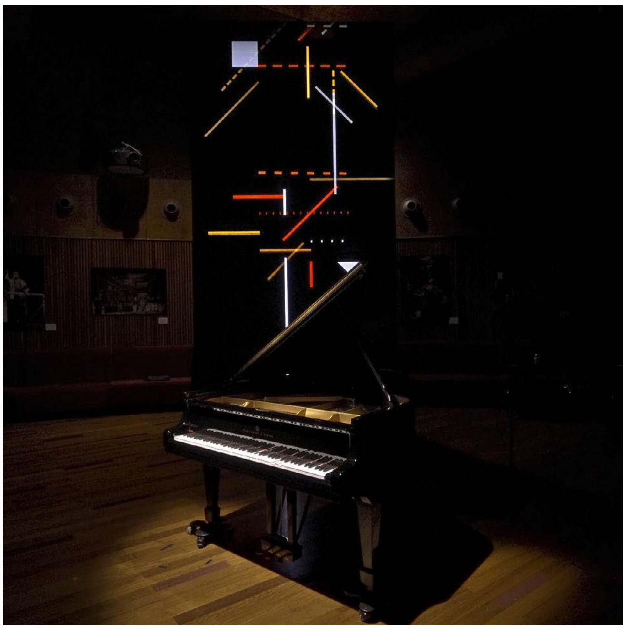


Figure 21. An image of the *ppff* setup in Barcelona's Liceu Opera. A vertical LED screen displays the downwards scrolling scores. When the graphics arrive at the piano position they are converted into MIDI to control a Steinway&Sons *Spirio* grand piano.

Musically, this solo piano piece comprises seven generative behaviors or miniatures. Each miniature delves into a distinct expressive, conceptual, or harmonic idea, ranging from atonal and indeterminate music to modal or rhythmic compositions which refer to and get inspiration from composers like Morton Feldman, György Ligeti or Arvo Pärt. All miniatures are heavily influenced by randomness, chance, and probabilities, ensuring that these musical pieces will always remain unique.

ppff remains a dynamic work-in-progress. Having been presented only once, it is set to undergo further evolution in the upcoming years. This progression will improve the image-to-MIDI methodologies and allow further research into the expansive potential of visual musical expressions inherent to such systems.

5. CONCLUSION

The FORMS initiative represents an advancement in the integration of contemporary generative music and digital art, particularly through its exploration of procedural graphic notation and Mel-spectrogram synthesis. Originating during the unique challenges of the 2020 pandemic lockdown, FORMS has demonstrated the adaptability and potential of software tools and algorithms in the realm of visual-to-auditory conversions.

This paper has provided a comprehensive examination of the various components and applications within the FORMS framework. From the foundational principles of image sonification to the diverse projects it has given rise, it is evident that FORMS seeks to address both the technical and artistic challenges inherent in this interdisciplinary field.

As the FORMS project progresses, it is anticipated to contribute further to the ongoing dialogue between visual art and music. Its continued development and refinement will likely offer valuable insights and methodologies for artists, researchers, and technologists aiming for a more cohesive and nuanced understanding of visual and auditory artistic expressions.

6. REFERENCES

[1] M. A. Boden and E. A. Edmonds, "What is generative art?" *Digital Creativity*, vol. 20, pp. 21–46, 2009. [Online]. Available: http://dx.doi.org/10.1080/14626260902867915

[2] K. Brougher *et al.*, *Visual Music: Synesthesia in Art and Music Since 1900*. London: Thames and Hudson, 2005.

[3] C. Roads and J. Strawn, *Graphic Sound Synthesis*. MIT press, 2003, vol. 32, no. 6.

[4] S. S. Stevens, J. Volkmann, and E. B. Newman, "A scale for the measurement of the psychological magnitude pitch," *Journal of the Acoustical Society of America*, vol. 8, no. 3, pp. 185–190, 1937.

[5] C. Roads, *The Computer Music Tutorial*. MIT Press, 1996, see sections on additive synthesis.

[6] G. T. Toussaint, "The euclidean algorithm generates traditional musical rhythms," in *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science*, 2005. [Online]. Available: http://cgm.cs.mcgill.ca/~godfried/publications/banff.pdf

[7] K. Perlin, "An image synthesizer," in *ACM SIGGRAPH Computer Graphics*, vol. 19, no. 3. ACM, 1985, pp. 287–296. [Online]. Available: https://dl.acm.org/doi/10.1145/325165.325247

[8] C. Reas and B. Fry. Processing webpage. [Online]. Available: https://processing.org

[9] Max-msp. [Online]. Available: https://cycling74.com/products/max

[10] Ars electronica festival. [Online]. Available: https://ars.electronica.art

[11] New art foundation. [Online]. Available: https://www.newartfoundation.art

[12] Institut Ramon Llull – Llengua i cultura catalanes. [Online]. Available: https://www.llull.cat/catala/home/index.cfm

[13] Hangar – Centre de producció i recerca d'arts visuals. [Online]. Available: https://hangar.org/ca

[14] Brossa Quartet de Corda. [Online]. Available: http://www.brossaquartet.com

[15] CosmoCaixa – Museu de la Ciència de Barcelona. [Online]. Available: https://cosmocaixa.org

[16] NEO – CosmoCaixa. [Online]. Available: https://cosmocaixa.org/ca/neo

[17] Oficina de Suport a la Iniciativa Cultural – Generalitat de Catalunya. [Online]. Available: https://cultura.gencat.cat/ca/tramits/oficina-de-suport-a-la-iniciativa-cultural/index.html

[18] Specifications – MIDI Association. [Online]. Available: https://midi.org/specifications

[19] openFrameworks. [Online]. Available: https://openframeworks.cc

[20] PlaymodesStudio, "ofxOceanode," https://github.com/PlaymodesStudio/ofxOceanode, 2023.

[21] Aggregate – Gamut Inc. [Online]. Available: https://gamutinc.org/aggregate

[22] Premiados – Laus Awards - ADG-FAD. [Online]. Available: https://www.adg-fad.org/es/laus/premiados

[23] Gran Teatre del Liceu – Barcelona. [Online]. Available: https://www.liceubarcelona.ca

EVALUATION OF DIGITAL MUSICIANSHIP IN HIGHER EDUCATION THROUGH PLAYING AND CREATING DIGITAL SCORES

Solomiya Moroz
University of Nottingham
solomiya.moroz@nottingham.ac.uk

Craig Vear
University of Nottingham
EMAIL LINK

Fabrizio Poltronieri
University of Nottingham
EMAIL LINK

ABSTRACT

This paper presents an evaluation of digital musicianship with digital scores. Primary data was gathered during the *DigiScore* 'Roadshow' of North American universities' music departments in 2023 and an extended research workshop in Avellino, Italy, 2023. The activities consist of interactive lectures and practice-based workshops that involve students by asking them to reflect on the nature of their digital musicianship through digital scores. In defining digital musicianship we adopt Hugill's definition of musicianship as "a person's ability to perceive, understand and create sonic experiences" [1], and expand upon this with Brown's *Sound Musicianship* [2]. Through the lectures and workshops, we gathered data using online interactive polls and questionnaires in person. This data gathering was divided into: a) skills; b) contexts, cultures & literacy; c) musical identity & creative practice; and d) perception & awareness of (digital) music. In the paper, we present an analysis of this data and correlate it with observations from these sessions. In our discussion, we point to the interconnectedness of the sub-categories and also draw a connection between digital musicianship and creativity.

1. INTRODUCTION

The Digital Score (*DigiScore*) is an ERC-funded research project that investigates the transformation of the music score through computational technologies. [1] Digital scores utilising computational technology and digital media are emerging worldwide as the next evolutionary stage in the concept of the music score (discussed in Vear's *The Digital Score* [3]). Emerging findings are suggesting that digital scores can generate new music experiences, innovative compositional approaches, novel performance opportunities, and broader accessibility for a vast number of musicians and music cultures around the world. The *DigiScore* project defines a digital score as "a communications interface of musical ideas between musicians utilising the creative potential of digital technology" [3]. While not a separate paradigm to the traditional paper-based platform of traditional/conventional scores, they can be considered the next evolutionary step in the development of the music score by advancing the creative potential of the medium that carries communicative properties of the musical idea. A critical feature of digital scores is the application of elements of rich media, animated shapes, interactive and haptic controllers and even the embodied movement of robots to convey the specific language of the music idea. Thus, digital scores can go beyond fixed notation, and they can embrace the affectual and phenomenological properties of the media. The primary aim of *DigiScore* project is to understand how digital scores shift creativity and musicianship.

[1] https://cordis.europa.eu/project/id/101002086

1.1 Background of "Roadshow" Research

A major work package in the *DigiScore* project is the "Roadshow." The aim of this is to evaluate higher education music students' wants and needs from digital musicianship education across the globe through engagement with digital scores. The method is to visit a diverse range of music departments across the world and work with the student bodies through interactive lectures and practice-based workshops with digital scores. A typical process would be to spend 3 hours with 4–6 students working through a selection of digital score types, followed by a 90-minute interactive lecture with the full cohort. At the end of this lecture we may perform the works that were workshopped with the students.

Parallel to the lectures and workshops, *DigiScore* developed extended workshops in digital score-making facilitated by a unique set of "creativity cards." [2] The first such workshop was delivered in Avellino, Italy between 11–14 October 2023. The research focus was on students' creative process while observing changes to their digital musicianship. Primarily, the creativity cards workshop supported students' development of a musical idea towards a realisation of a functioning prototype of a digital score. This offered us the potential to study any transformational changes in creativity that can take place through digital score making through questioning, observations and consultations with their academic professor.

2. METHODS AND FINDINGS

2.1 Aims of Digital Musicianship Research

The North American tour (February 2023) gathered data on digital musicianship from the point of view of "a per-

[2] Further information can be found at https://digiscore.github.io/pages/Creativity_cards_about

son's ability to perceive, understand and create sonic experiences" [1]. The principle aims of evaluating digital score musicianship during our presentations and workshops were focused on the following criteria:

- **Skills**: what are the skills needed to articulate and interpret features and effects of digital score musicking?
- **Contexts, Cultures & Literacy**: what contextual, cultural literatures and insights are required to inspire creative thought and support musicking ideas?
- **Musical Identity and Creative Practice**: what are the new modes and possibilities of creative practice?
- **Perception and awareness of (digital) music**: how do musicians actively analyse digital score music, and what interpretations are they generating when making music?

2.2 Methods

In the case of the North American tour, we used three types of methods to evaluate musicianship with digital scores. These were:

1. A Mentimeter poll [3] which allowed students to contribute responses through a 90-minute presentation and online survey.
2. After the lecture, there was an online survey questionnaire where students could answer in more depth the questions asked during the lecture.
3. For students participating in practice-based performance workshops with digital scores, there was a questionnaire aimed at evaluation of their creativity, transformational potential and students' wants and needs from higher education.

2.3 Lecture Presentation Findings

Throughout the 4 weeks of the North American tour, we visited public and private universities with music departments that ranged from large internationally renowned centres for innovating electronic music (such as Columbia and Carnegie-Mellon), or music performance and composition schools (such as Illinois at Urbana-Champaign, Northeastern and New England Conservatory), to small departments with a focus on pop music production and music business (such as New Haven). We spoke with undergraduate, and postgraduate students and PhD researchers, postdocs and faculty members.

Overall, 92 students engaged with the in-lecture Mentimeter polls, from which 50 completed the online form. The findings from both revealed a few tentative trends in digital musicianship. For example, when asked how students would define themselves on a poll with multiple answers such as composer, improviser, performer, electronic musician, instrument maker, etc., 38 out of 92 students (41%) identified themselves as both composers and performers in addition to other multidisciplinary categories (Fig. 1).

[3] https://www.mentimeter.com

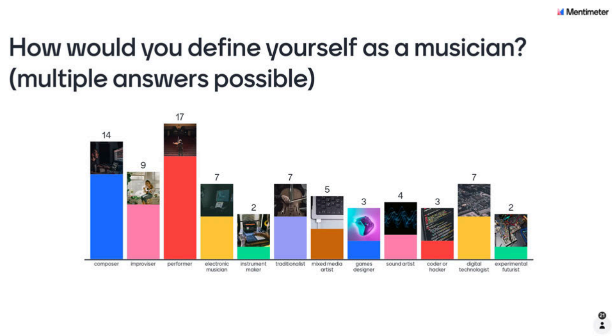


Figure 1. 1st slide from the *DigiScore* roadshow questioning strategy from the Mentimeter poll.

In the online survey, we also found that 66% of students were seeking a variety of digital music skills, including coding, basic audio engineering, recording and mixing (DAW), analogue circuitry, etc. (Fig. 2). These students were already media-curious and used a range of tools and disciplines in their music-making. One student mentioned their skills as: "Classical training on piano and violin, improvisation, coding in Max/MSP, training as a composer, training as a dancer" (Anonymous, 2023).

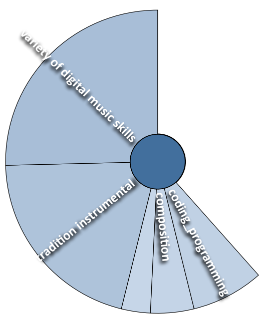


Figure 2. 33 (66%) coding references for *variety of digital music skills* pursued in Digital Skills category of online questionnaire.

Our findings also pointed to a heightened awareness of the need for coding and AI skills in the current and future workplace. Through the Mentimeter poll, we found that 37% were interested in actively pursuing machine learning and AI skills, and developing core proficiency's to prepare for this need, for example:

> I play a couple of instruments at a beginner to intermediate level, comfortable with multiple DAWs and many pluggins, and am proficient at multiple coding languages including python, C++, C#, HTML, and learning a few others. (Anonymous, 2023)

In the online surveys, we found that the skills that musicians were seeking were contributing to how they viewed their digital music identity. 42% mentioned that digital technology was extending or enhancing their music-making. This in turn made their music "more diverse" contributing to "more efficiency" in their workflow, "allowing a broader sound palette to work with and more access to different sounds" (Anonymous, 2023). As the result of working with digital tools and approaches, some students described their music identity as "creative technologist," "producer/instrumentalist," "improviser and electronic musician," and "composer and producer," which aligns with Mentimeter poll results (Fig. 1).

Overall, students mentioned technology as an important tool in their creative practice of making music. In the *musical identity and creative practice* theme, one student mentioned: "A common thing I practice is the translation of electronic techniques to my instrument" (Anonymous, 2023). Furthermore, there was a general awareness of how digital technology shapes one's musicianship through their creative practice, for example: "I am an explorer of what is out there. The tools change the affordances and therefore my result. I am constantly exploring the variety of tools so I may benefit from them all" (Anonymous, 2023).

When we asked about what might inspire their music in a Mentimeter poll, 64% sighted other media as inspiration. This was a broad litmus test asking them to choose one or more of the following:

- Types of music/musicians
- Books/theories
- Art forms (art, theatre, dance, etc.)
- Media (film, gaming, video, etc.)
- Nature and the environment

Overall, "types of music/musicians" were the dominant choice, but not by much. Generally, only peeking over the top of others by 1 or 2 counts. With the others generally coming out equal. Here, one conclusion is that students are demonstrating aspects of pluralism by taking inspiration from many more sources than merely music studies.

In support of the emerging idea of a pluralistic music identity, we also found that students were engaged in different styles, genres and musical concepts in their music-making. We found that 58% (Fig. 3) draw on different cultural influences and this reflects the digital skills that they are pursuing. While 66% had, or were pursuing, digital music skills, and were from pluralistic backgrounds. For example, one student mentioned the music context that they are working in and the music that they are inspired by: "Electronic music, free improvisation, contemporary classical music, sound art and installation. I'm inspired by traditional Mexican music: son jarocho, son huasteco, norteño, cumbia, reggaeton, sonidero" (Anonymous, 2023). Others mentioned an interest to learn music from other cultures: "I'm mostly interested in working in improvised, electronic, and experimental; however, I have interest in other cultural music from other communities. I love to learn about the musical traditions of other cultures" (Anonymous, 2023).

In online questionnaires, when it came to the *perception and awareness of (digital) music* theme, we found that students often described what they value most in their music-making from the point of view of connection and communication. One student wrote: "I like how I can showcase my music to other people, as it gives me inspiration for what I can create and accomplish," and "for me, getting the right mood across to the listeners is a big part, making sure I'm able to make an impact within the song" (Anonymous, 2023). Similarly in the Mentimeter poll, when we asked the same question, a lot of students answered with themes such as "emotion," "community," "connection with others," "connecting with emotion," "sharing with others," "my creative voice," and "honesty."

The "Roadshow" lecture findings point to multidisciplinary identity in music making, inspiration by other media and not just music and other musicians, the pursuit of AI and machine learning skills, and pluralistic music identity. We are also finding that our four themes of *digital music skills*, *context, cultures and literacy, musical identity and creativity, perception and awareness of (digital) music* are mutually interdependent. For example, digital music skills that one is pursuing are often linked to the context in which one may be creating their music as well as the musical identity and creative processes that they are engaging with. Furthermore, it is through the perception and awareness of their digital skills and how they are perceived in a context with other musicians and audiences that musicians learn about their musical practices and their (digital) music identity.

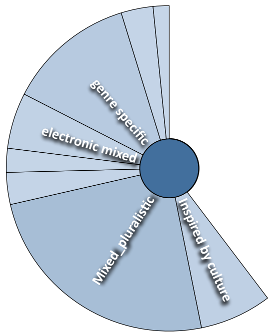


Figure 3. 31 (58%) coding references for *mixed pluralistic cultural influences* in Contexts, Cultures & Literacy category of online questionnaire.

3. DISCUSSION OF LECTURE FINDINGS

The concept of the mutual interdependence among *skills, context and cultures, musical identity and creativity, knowledge and awareness* suggests that these elements are not isolated but rather function dynamically aligning with dynamic system's theory (DST) [4]. In this framework, creativity plays a pivotal role in the ongoing process of learning and expressing one's musical ideas. It draws inspiration from an individual's context and background, contributes significantly to the formation of musical identity, and aids in the self-evaluation of digital skills and learning processes.

Our perspective on creativity differs from viewing it as a product solely emanating from creative individuals. Instead, we consider it a phenomenon that emerges through the interactive dynamics among multiple agents and their connection with the broader socio-cultural environment in which they operate [4]. This viewpoint allows us to better comprehend the dynamic processes inherent in the diverse activities and approaches that current higher education (HE) students undertake as part of their digital musicianship. These activities often unfold in collaborative, self-directed learning environments, extending beyond the structured formal education setting. One student noted: "I value collaboration a lot within my music-making. My most important aspects are the instrumentals & vocals" (Anonymous, 2023), and "my main focus is hardcore crossover due to my band, but I enjoy playing jazz, funk and other types of rock in my freetime" (Anonymous, 2023). These expressions of musical preferences and engagement highlight the dynamic interactivity at play, creating feedback loops through embodied and socially situated practice [4].

In essence, the interconnectedness of skills, context, cultures, musical identity, and creativity is not just a theoretical construct; it manifests in the real-world activities of contemporary HE students involved in digital musicianship. The dynamics unfold within collaborative, self-directed learning environments, emphasising the rich, interactive nature of the processes involved.

To draw the point of the interconnectedness of these categories as facilitated by creativity as a process, we would like to discuss further the dynamic system's approach. In our responses, we found that 26% of students came from specific cultural music environments or identified their creativity to be embedded within them. For example, someone who considers their cultural background and literacy within rnb, jazz and pop might be interested in acquiring skills that are aligned with some traditional music skills such as performing, composing recording and mixing in DAW but also learning about processed and autogenerated sound that could extend their jazz practice to "create a more modern sound that would not have been heard before" (Anonymous, 2023).

58% of all respondents draw on different cultural influences and this reflects the types of digital skills that they are pursuing and the music that they are making. For example, one student mentioned a mix of electronic, improvised, sound art with traditional Mexican, cumbia, reggaeton, and sonidero influences on their music-making while identifying with electronic and improvised music. The fact that students are both embedded and seek different cultural influences speaks to the socio-dynamic aspect of creativity that they are involved with. But equally 42% did not clearly indicate that they were drawing on different cultural influences. This property seems quite balanced, and given the broad flavour of the universities that we visited, probably expected. Interestingly, nonetheless, to note that, around 50% of these students that we polled are drawing on influences that are extra-musical and beyond the typical focus of the academic music curriculum (or at least the ones us authors are aware of).

66% of all respondents had or were pursuing a variety of digital musical skills and were also from a pluralistic background. Here we can consider dynamic processes of co-determination that occur between musicians when they interact in music settings which are mixed between different categories of skills, culture and identity, music identity and creativity, and knowledge and perception. Interestingly, 18% of the participants valued knowledge and awareness of their musicianship within the context or genre within which they were creating. This demonstrates another element of learning and development that is supported through a dynamic system theory approach, whereby critical evaluation is part of what musicians do to see how new or original a certain approach or use of a digital skill could be in their practice. Andrew Hugill cites this as "a curiosity, questioning and critical engagement about what they (digital musicians) do" [1].

4. PRACTICAL WORKSHOPS

To discuss further creativity specifically with digital scores it is interesting to view the responses from the surveys from both the practice-based performance workshops and the creativity cards workshops. Both sets of workshops gathered a smaller amount of participants' answers, 6 surveys collected from the North American tour and 7 from the creativity cards workshop in Avellino, Italy.

4.1 Performance Workshops

In the practice-based performance workshops, we presented students with a range of digital scores which included digitised, animated, interactive, system digital scores, and intelligent. This was essential to show students the full continuum of possibilities of what a digital score could be as discussed in *The Digital Score* [3]. The aim of the workshop was to support the students' ability to interpret and engage with the range of the digital scores in order to deliver a short digital score performance for their peers at the end of the session.

In the post-workshop questionnaire, we asked students about aspects of creativity with digital scores, their transformational potential and their wants and needs from formal education to engage further with digital score performance. In our surveys, most students remarked on the heightened engagement experienced with the media of the digital score, compared to the traditional score: "My engagement with the scores was more visceral than it may have been were I looking at a score on a piece of paper," while another emphasised, the "animated nature of many of the scores draws my attention to exactly where it needs to be in the score at any given time" (Anonymous, 2023).

Students also appreciated the openness of the digital score, exemplified by pieces such as Gudmundur Steinn Gunnarsson's Quartet [5], where musicians can choose instruments or pitches to interpret with a rhythmic notation on a scrolling timeline. One student expressed: "Pieces like the Gunnarsson's quartet or John Cage's *Variation* provided firm enough ground rules that offered a vast amount of freedom, while not feeling overwhelmed," and another stated, "digital scores seemed to present a source beyond just a composer's outline for a piece" (Anonymous, 2023). With [5], they appreciated the simplification of complex rhythmic notation, enabling them to engage with an animated timeline instead of counting or observing different time signature changes. “My greatest joy in this piece is that I didn’t have to count, I could just play, which is rare in rhythmically-complex scores,” and “it is more accessible than seeing different time signature changes on a page” (Anonymous, 2023). The animated aspect of the graphical interface in these scores had a liberating effect on students’ engagement and presence while performing. In addition, students remarked on the ease with which one’s attention can be drawn towards making abstract relations between sound and what is seen on screen: “I am never at a loss as to where our attention is being directed in the score, and as a result, it is very easy to draw abstract relations between sound and what is on screen” (Anonymous, 2023).

The interactive playful nature of some digital scores such as *Nautilus* [6] and *Plurality Spring* [7] led to spontaneous interpretation for some students that they found would not be possible with paper scores. Both of these digital scores represent interactive game engine digital scores where one’s attention is drawn to emerging aspects of the score stimulated by the sounds of the musicians’ instruments. Musicians found the interactive and gamified nature of the notation provided numerous possibilities for rhythmic interpretation [8]. The game-like behaviour of some of the digital scores offered positive feedback to participants and encouraged an impulse to ‘win’ or play a score like a video game, changing the understanding of performing the ‘right’ actions as one would in a performance: “I felt like I could ‘win’ by performing the right action, and this impulse took over. At this point, I was no longer performing – I was playing how someone might play a game” (Anonymous, 2023). The gamified approach expanded rhythmic possibilities, as one student confirmed: “the rhythmic possibilities of doing a ’game’ piece seem endless” (Anonymous, 2023), sparking creativity in the interpretation of the digital scores.

The animated and gamified nature of digital scores was very engaging for some students, taking away any anxiety associated with sight reading a traditional piece of music: “The digital score version not only stripped the reading of any anxiety, it turned what might be a very tediously-prepared piece into a fun game” (Anonymous, 2023). Moreover, this gamified aspect of digital scores helps us to understand their appeal, as gamification can be understood as a “general process in which games and playful experiences are understood as essential components of society and culture” [8]. Students felt immersed in the moment of music making, particularly when interpreting scores that have been transformed from paper notation like John Cage’s *Variation I* to an animated digital format using Decibel ScorePlayer app [9]. Workshop participants recognised and appreciated that the composers were exploring new ways of engaging with performers through their digital scores.

There were also challenges experienced by the students. One significant issue that was managed by the *DigiScore* team, was the amount of contextual explanation that was necessary before the student musicians were able to interpret the score successfully. By that we mean, that they did not have the necessary conceptual training of interpreting animated marks, colours, symbols into a notation that they could realise. And that each new digital score type required a new type of conceptualisation. However, once this had been carefully explained and contextualised they felt a little more sure about how to interpret the music. This was also due to the way the team explained these concepts from the perspective of musicking, and relating abstract media-concepts into usable information for use inside music-making.

Through this, and perhaps because of the careful explanations, the students recognised the advantages of working with digital scores in higher education. They expressed a desire for more accessibility to digital scores, including a repository, easy packaging of digital scores and learning tools for diverse music genres. Overall, they felt that digital scores could foster certain students’ motivation and creativity in music education.

4.2 Composition Workshop With Creativity Cards

For the first half of 2023, the *DigiScore* principal investigator designed and developed a creativity card resource for use in digital score practice-based workshops. These cards aim to scaffold the design, development and creation of digital scores. They present insights from different perspectives about the possibilities, opportunities, challenges and questions around creating a digital score. They have been designed to operate like a conversation between workshop participants, their creative ideas, and the experience from those who have researched and specialised in making digital scores.

[10] points out that decks of cards are a long-established tool to aid design, fostering an environment to encourage creativity through play and problem-solving. In music, *Obliques Strategies cards*, pioneered by Brian Eno and Peter Schmidt, have been used in suggesting different paths through a repository of tools for artists and musicians to overcome creative blocks [10]. Eno clarifies that the cards’ role is to provide a broader perspective when it comes to finding creative solutions: “The cards evolved from me being in a number of working situations when the panic of the situation [...] tended to make me quickly forget that there were other ways of working and that there were tangential ways of attacking problems that were [...] more interesting than the direct head-on approach” [11]. Over the past decades, a variety of cards have been produced aimed at stimulating creativity in fields such as marketing, management as well as design, especially user experience (UX) and digital design. Like in the cards for design, the *DigiScore* creativity cards are also not prescriptive but act as a support for inspiration, organization and communication of ideas of the digital score [12]. Thus, it is no surprise that they could be used in an intermediate endeavour like making a digital score, and like in UX design, they could be aimed at creating an experience inside the music-making process for the musicians.

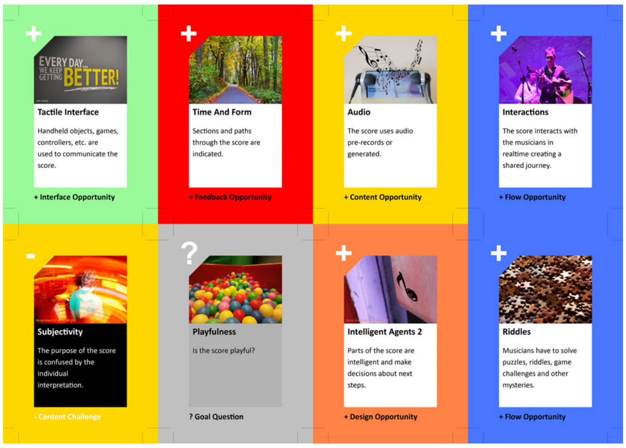


Figure 4. Print sheet from the *DigiScore* creativity cards collection, showing 8 different types of cards.

The direct inspiration for the *DigiScore* creativity cards was Wetzel et al's *Mixed Reality Game* (MRG) cards [13]. On the history of ideation cards, they state:

> The physical properties of ideation cards make them resemble card games, and they can be classified as design games [14] [15]. Ideation cards possess game-like rules ranging from the way they structure card draws, turn-taking, playing and discarding cards to randomly revealing them. They enable collaborative design in a playful atmosphere. Cards are used as orienting devices, conversation starters, and pace-makers [16]

[17] describes ideation as "a matter of generating, developing and communicating ideas, where 'idea' is understood as a basic element of thought that can be either visual, concrete or abstract." As physical artefacts, the cards allow groups to think through the process of composition, a type of exploration in which the participants "play around with materials without knowing what will come of it" [18].

Like Wetzel et al's MRG cards, the *DigiScore* creativity cards are designed for two purposes:

- First, to open out creative thinking by presenting possibilities and systems for idea generation within the context of digital score creativity.
- Secondly, to present known problems, challenges, and opportunities as springboards for design selection or refinement.

Currently, the *DigiScore* pack comprises 152 cards split into 3 different sections: *opportunities, challenges* and *questions*. Each of these sections is split into 7 different modes of thinking that align with the 7 modalities of a digital score [3]: *interface, design, content, language, goal, feedback, flow* (Fig. 4).

The first creativity cards workshop was presented at Cimarosa Conservatory in Avellino, Italy, October 11–14, 2023. In the 3-day workshop, we worked with two groups of three and four students respectfully. They were mostly first-year music composition students with an addition of a couple of older year students. Throughout this process, the research team also observed students' creative engagement in the workshops and delivered surveys tracking how creativity cards contributed to students' digital musicianship.

4.2.1 Observing Creativity, 1. Communications Skills

In observing students' activities in the workshop, we employed two existing theoretical frameworks to identify communication and creative engagement with the cards. The first one focused on their communication through a framework known as COGS (Communication Opportunity Group Scheme) [19]. This supports observations of an individual's progression/development in communication skills through the activity. These skills are:

- Personal qualities
- Social qualities
- Decision-making
- Communication (verbal and written)
- Performance (showing, demonstrating, presenting)
- Interaction with technology

To observe these skills in action we trialled a course grain approach of logging specific students growth over three steps: passive, engaged, and leading. The aim was to capture any indication that either the cards or the engagement with digital scores could be enhancing these extra-musical skills through the workshop process. This task was undertaken by the research assistant of the project using a simple matrix sheet of skill vs development for each target student.

Over the course of the workshop it became evident that this method was not going to work as originally planned due to the dynamic nature of interactions in the group. As such, these categories were followed only loosely; thus, our evaluation of their activities was not judged individually but from the point of view of a group activity. Also, students were interacting mostly in Italian, and our full ability to judge their overall performance in these categories was limited. However, one core member of our group did speak fluent Italian and was able to report on the verbal communication aspect of this study. The advantage of the COGS framework is that growth in communication skills beyond text and spoken word can still be observed through the students' engagement with the task.

Throughout the workshop, we noticed some students' progress in their communication skills and understanding of the task at hand. There were observable elevations in all categories of COGS across these students. Overall, we found students were responsive to questions posed by the *DigiScore* team, were highly motivated and engaged in the activities of decision-making and creation with creativity cards.

Of the evidence that we gathered, it is difficult to state what caused this seemingly elevated state of engagement. We do not know if this was because of the activity, the cards or engaging with digital scores, or whether this was their natural behaviours and personal attributes emerging through the workshop as they became confident or comfortable with the *Digiscore* team and the workshop. Consultations with the students' professor provided some evidence that both were true, but we have no quantitative evidence to support this beyond his testimony.

In the future, this essential aspect of the research will need to be enhanced to include bookend tests of the students' baseline attributes in communications. The COGS scheme does supply such a test, and this has been proven effective across numerous trials. It was not implemented in this workshop because we wished to test the relevancy of such an approach. So, from this perspective, we feel that COGS is a crucial method with which to study the impact of digital score engagement on the communication skills of music students, and will be employed in future workshops.

4.2.2 Observing Creativity, 2. Possibility Thinking

The second theoretical framework examined students' creative engagement with the cards through Possibility Thinking (PT) [20]. PT is mostly used in pedagogical environments in the context of children's learning through the following categories:

- Posing questions
- Play
- Immersion and making connections
- Being Imaginative
- Innovation
- Risk-taking
- Self-determination

We assessed students' engagement through questions, and observing their immersion and connection to the task of making a digital score, being imaginative in realising a prototype digital score idea, taking risks with the idea of making a digital score as no one in our workshop had done this before, and a degree of innovation with realising complex technical music ideas with simple means. Overall, we find PT was useful in observing creativity through the creativity cards workshop.

Another important aspect of PT is *play*, and it was observed in students as they enacted musical activities, playing through and with creativity cards and trying their digital score ideas with musical instruments. We also observed a relational approach to making music with the help of creativity cards, whereby students interacted with each other through collaboration, communication, self-organisation and playing with ideas. This speaks to the embodied approach to learning whereby students were adapting through dynamic interactivity in relation to each other's activity [21].

In the post-performance discussion, students said that creativity cards presented them with "many choices, opportunities and possibilities for making a digital score" (said in Italian and translated by the course teacher). Students also appreciated that the focus of creativity cards stayed on their embodied experience within the digital scores that they were making, being immersed in musical situations rather than technical or programming ones (observed in the post-performance discussion). It was interesting to note that one of the students mentioned a "destabilising effect" (said in English during the post-performance discussion) of creativity cards as it introduced a new approach to composition that she had not encountered before. However, she also admitted that "destabilisation" was productive for the digital score idea as it introduced a new way of thinking and resolving problems. The destabilising effect can also be viewed from the point of view of DST, whereby a new action may introduce entropy into an organism or, in our case, a musical situation that would need to be resolved contributing to learning a new skill or a creative approach [4].

Structuring the workshop in this way and allowing new ideas to emerge within a group dynamic has many challenges. While the activities were structured using the creativity card format and programme [4] and managed by the experienced team, it led to many heated discussions amongst the student groups, one in particular being overly vocal about it. This was more noticeable during the initial brain-storming activities were minds were allowed to run wild in the 'candy-store' of possibility. It is a useful activity, but on the other hand, it needed careful management so as not to detract from the main focus of the workshop which is building a rapid prototype version of their idea. During this phase, the teamwork became more apparent, but with a lot of guidance and management from the experienced team.

With the final presentation on day 3, we observed that while the original ideas for digital scores in both groups were very ambitious, incorporating sensors, AI and complex sound synthesis, both groups were able to reduce the technological demands of their digital scores through creativity cards. The fourth round of cards asked them to reconsider the ideas brainstormed so far pushing towards a finer focus for the performance. In addition, the embodied and practical way of working with the digital scores enabled students to test through their playing what could work to realise a complex idea with the tools and means available to them in the moment of making. In the end, students reflected positively on their digital score prototype ideas. They appreciated that creativity cards enabled them to generate many ideas giving opportunities to make a digital score faster, promoting fast thinking and doing. This type of process prioritised music-making and not just programming, starting from the practical aspect of making a digital score with limited means possible. In this way, technology became a tool for students like a piece of paper in the traditional composition approach.

4.2.3 Digital Musicianship Surveys and Creativity Cards

To look further into what students experienced through the 3-day workshop, we can turn to the digital musicianship surveys that were captured each day of the workshop.

The digital musicianship surveys focused on the four areas of digital musicianship as mentioned in 2.1. However now, these categories were measured progressively throughout the 3-day workshop. The aim was to notice any changes in *digital music skills*, *contexts and cultures*, *digital identity and creativity* and *knowledge and perception* over time. This was structured as:

[4] introduced at https://digiscore.github.io/pages/cards_how_to_use

- An initial questionnaire on students' musical background was taken enquiring on their skills, how they normally present and communicate their music, how they describe their musical identity and their process of making music, and what they value most in their music making.
- At the end of day 2, the second survey focused on whether there was an extension to the musical skills that students previously had, new approaches to composition or did the students use the skills already at their disposal. The second question in the survey asked whether making a digital score changed the way they normally communicate music, while the third question focused on how making a digital score changes the way musicians perceived themselves and whether it extended their musical process or introduced something completely new; while the fourth question focused on what musicians perceived as interesting and useful about working with creativity cards to make music.
- The third survey enquired whether, through the overall process of making digital scores, students had acquired new music skills and what new aspects of music-making students would like to keep for their future work. The questionnaire also asked whether the final presentation reflected well the digital score idea that the students were working on with the creativity cards and, given the experience of this workshop, will it influence how students present their future projects.

In the category of *digital musicianship skills*, we had a very diverse group of students, some using more established digital music skills such as digital audio workstation (DAW) programming, and others just starting to compose instrumental music. For example, 3 out of 7 students (43%) mentioned DAW skills and working with sound analysis tools. Students with digital music skills mentioned building on their previous skills and in some instances acquiring new skills such as making a digital score or playing music with instrumentalists as previously they only made electronic music.

Students who did not have previous digital music skills also learned about new types of software that could be used in digital score making in addition to being introduced to a completely new way of composing. Some of the answers related to the digital skills used and learned in the workshop: "I used skills learned over the years and learned new ways of writing and playing music," and "they were an extension of what I knew, a new way of composing" (Anonymous, 2023). Others remarked on the novelty of the workshop as: "New, I would never have thought of such an approach," and "I relied little on what I knew as it was all new information" (Anonymous, 2023). Another student remarked that the new skill as facilitated by creativity cards was: "Knowledge of what a "digital score" is and working in the shortest time possible for maximum yield" (Anonymous, 2023). Here, we saw progress in the digital music skills that students acquired through the workshops. All students seemed to have acquired new skills, either through working with other students who already had digital music skills or being introduced to what a digital score is and how to make one in the shortest amount of time possible.

In *contexts and culture*, most students mentioned their background in contemporary/classical music. 3 out of 7 students (43%) already were fluent in electronic and digital music and the context in which they presented their music reflects this. It was interesting to observe through the survey answers on days 2 and 3 that most students felt the workshop changed their perspective on the way music can be made as one student mentioned:

> Creating a score different from the one usually used opened my mind to new ways of expressing and perceiving music by listening and welcoming what other musicians do. (Anonymous, 2023)

And how it may inspire their future music making: "Surely, it's a way of making music that will certainly influence my future projects," and what they learned through the creativity cards workshop:

> A greater curiosity in drawing from the technological world a series of tools that can expand the language. The understanding of a musical context that can no longer be divided into watertight compartments: composer, performer, audience. (Anonymous, 2023)

Most students understood the freedom implicit in making and performing digital scores, the change in aesthetic perspective as there is no one composer/performer but all are equally involved, and the openness of possibilities to explore further in the digital tools and ways of making music.

In *digital identity and creativity*, most respondents already mentioned that their musical identity and process were generally very open to new influences, contexts or circumstances of making music. This we do not find surprising as most of the students were in the first year of studying composition. However, even at this stage, they found creativity cards' way of working transformative to their music making. Some mentioned that creativity cards extended their vision when it came to collaboration with others. While others said: "It introduces the ability to not abandon ideas, be more practical and use more instinct" (Anonymous, 2023). One student found it an additional tool to their composition work: "[I] think it is a parallel path to traditional composition work, certainly very interesting" (Anonymous, 2023). Some found it very transformative to their music identity and future creativity: "I am no longer a defined identity (director, composer or performer) but I am a musician who mixes with other musicians by processing their inputs" (Anonymous, 2023). Some other transformative experiences of music making were stated as: "The experience changes the way of interacting with the score and, therefore, with the music, extending musical discourse to new vocabulary" (Anonymous, 2023), and "it is extending my way of making music and this can only make me a complete musician" (Anonymous, 2023).

In *knowledge and awareness*, reflecting on the usefulness of creativity cards for their digital score idea during the workshop: “The most interesting thing was starting from the core of the idea and then expanding it,” and that in future work, the same participant said: “I will be more careful to integrate different aspects of music, not just electronic” (Anonymous, 2023). Students mentioned how the creativity cards workshop was useful in making digital score prototypes: “To reach a working goal in the shortest time possible” and “the most interesting thing was the collaboration” (Anonymous, 2023).

Other statements on how working with digital score creativity cards was transformative and impactful: “The openness to new writing systems that include technological tools not “only” as tools, but as a source of,” and “the fact that making music is not just playing an instrument. It helps me to be aware of the fact that everything around us can be a composition” (Anonymous, 2023). Students also reflected on how creativity cards’ approach to composition might influence their future compositions: “I will modify them in an innovative way that is different from what I have done until now,” and “this project will certainly have an impact on my vision of making and creating music” (Anonymous, 2023). Other transformative statements were: “It’s definitely a mind-changing experience that will influence my way of playing, composing and just generally creating,” and “there will be greater attention to the roles involved in a performance, to the importance of a score that can be generated live, to an interactivity that develops together with others” (Anonymous, 2023).

4.3 Discussion of Creativity Cards

Through the digital musicianship surveys, we can track clear changes that took place in students’ musicianship. They indicate that through an engagement with creativity cards these students had new and transformative experiences in making and performing music. When it came to digital music skills, creativity cards either extended musicians’ previous skills or introduced completely new tools and ways of making and performing music. In the post-workshop reflective report, one student mentioned that “we had the opportunity to discover a new way of making music that goes beyond what we are used to” (Anonymous, 2023). We also noticed positive changes in the way musicians viewed themselves and the kinds of activities they would want to engage in the future as a result of working with creativity cards. 4 out of 7 students (57%) mentioned the interchangeability between ‘composer’/‘performer’ in the making and performing digital scores as truly transformational to how they will view music-making in the future. These students also appreciated the accessibility of digital scores that could invite untrained musicians or the audience into making music.

In analysing the evolution in these digital musicianship surveys, we can see that the changes that took place in students’ musicianship could be supported by our observations from a three-day workshop on students’ creative process. Here, we observed students’ engagement, immersion and play with digital creativity cards, where we can conclude that meaningful interaction took place. Furthermore, these interactions took place within group dynamics whereby students posed questions and came up with innovative ideas for their prototyped digital scores through lively and embodied interactions with each other, demonstrating that both 4E creativity [4] and PT [20] took place. Additionally, meaning was made in relationships that students built with the score materials and each other during the workshop, as many mentioned “collaboration” and “communication” with the score and each other that formed part of their reflection on whether they would use this approach in the future.

There were some other conclusions that are worth noting here, although there is less conclusive evidence to support them. First, it is interesting to note that the creativity cards introduced a destabilising effect for composing music. Here, we see no surprise since the idea of using cards, although has been used in music, comes from the field of graphic design, and none of the students have engaged with this way of creativity before. However, the novelty of introducing creativity cards to making music indicated an opportunity for divergent thinking which in itself is an endeavour resulting in higher creative gains [21]. Another point worth noting is that these cards have the potential to bring the benefits outlined by [22], as they allowed the framing of problems from multiple perspectives, leading to externalisation of insight to facilitate dialogue while providing a way to use existing knowledge that may have been dormant. The cards also offer frameworks for visualising problems and solutions and equip teams for learning about people’s experiences. And finally, the cards offer a type of creative engagement involving both combinatorial and divergent thinking, pushing students to collaborate with each other and experience their digital scores from an embodied perspective.

5. CONCLUSIONS

In this paper, we presented the findings from two separate, yet interlinked, activities: a “roadshow” of North American university music departments that included interactive lectures and practice-based workshops, and an extended three-day practice-based workshop in Italy. The focus was on engaging student musicians with digital scores in order to gain insights into their perspectives around digital musicianship and their needs and wants from contemporary music education. We surveyed roughly 60 students and engaged 12 in the practice-based workshops. While this is not an exhaustive survey, our findings point to emerging trends in students reflective understanding of musicianship through these four themes: 1) *skills*; 2) *context*, *culture and literacy*; 3) *music identity*; and 4) *creativity, knowledge and perception*.

We speculate from these findings that these themes are interlinked and dynamic with each other. Furthermore, that this dynamic interrelationship is often a socially situated, interactive and embodied experience for these musicians. In studying the dynamic behaviour of these four themes, it has been shown that many musicians in our study possess a pluralistic music identity that does not reflect one genre or cultural background, and this extends their digital music practice. From our practice-based workshops in performance and composition of digital scores, we observe and conclude that creativity as a dynamic process is shaping musicians' meaning-making with digital scores which further transforms their digital musicianship.

However, we note that this is not self-evident. The students that we engaged with needed concepts explained to them, workshops managed, and the translation of extra-musical or media theories into a language that is relevant to the domain of music. Furthermore, the process of conducting practice-based experiments with new concepts such as gaming, animated scores, AI and robotics opened up the 'cookie jar' of possibility and potential that led to heightened states of [over] excitement. We might conclude that if, as music educators, we are to adopt emerging insights into the nature of current digital musicianship, and the wants and needs of some of the students in our HE system such as those presented here, we may find it necessary to update our current ways of thinking and adjust how we instruct digital musicianship.

Acknowledgments

The Digital Score project (*DigiScore*) is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. ERC-2020-COG – 101002086).

6. REFERENCES

[1] A. Hugill, *The Digital Musician*. Routledge, 2018.

[2] A. Brown, *Sound Musicianship: Understanding the Crafts of Music*. Cambridge Scholars Publishing, 2012.

[3] C. Vear, *The Digital Score: Musicianship, Creativity and Innovation*. New York: Routledge, 2019.

[4] D. Van Der Schyff, A. Schiavio, A. Walton, V. Velardo, and A. Chemero, "Musical creativity and the embodied mind: Exploring the possibilities of 4e cognition and dynamical systems theory," *Music & Science*, vol. 1, 2018.

[5] G. Steinn Gunnarsson, "Kvartett no. 5, 2nd movement (2010) [w/ screen score]," https://www.youtube.com/watch?v=lzWhcOYOKcU, accessed Oct 25, 2023.

[6] "Vear, C. DigiScore: Nautilus (2022). Performed by Carla Rees." https://www.youtube.com/watch?v=XK-9eXCJxCg, accessed Oct 25, 2023.

[7] S. Hutchinson and P. Turowski, "Plurality spring (performance reel)," https://www.youtube.com/watch?v=2JhDLg07_a8, accessed Oct 25, 2023.

[8] M. Fuchs, S. Fizek, P. Ruffino, and N. Schrape, *Rethinking Gamification*. meson press, 2014.

[9] "Decibel scoreplayer. an application for ipad," https://decibelnewmusic.com/decibel-scoreplayer/, accessed Nov 15, 2023.

[10] R. Roy and J. P. Warren, "Card-based design tools: A review and analysis of 155 card decks for designers and designing," *Design Studies*, vol. 63, pp. 125–154, 2019.

[11] C. Amirkhanian and B. Eno, "Brian eno interviewed on kpfa's ode to gravity," 1980.

[12] S. Baldwin, "Ux ideas in the cards," *UX Magazine*, vol. 3, p. 615, 2011.

[13] R. Wetzel, T. Rodden, and S. Benford, "Developing ideation cards for mixed reality game design," *Transactions of the Digital Games Research Association*, vol. 3, no. 2, 2017.

[14] E. Brandt and J. Messeter, "Facilitating collaboration through design games," in *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices-Volume 1*, 2004, pp. 121–131.

[15] K. Halskov and P. Dalsgård, "Inspiration card workshops," in *Proceedings of the 6th conference on Designing Interactive systems*, 2006, pp. 2–11.

[16] E. Hornecker, "Creative idea exploration within the structure of a guiding framework: the card brainstorming game," in *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, 2010, pp. 101–108.

[17] B. Jonson, "Design ideation: the conceptual sketch in the digital age," *Design Studies*, vol. 26, no. 6, pp. 613–624, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142694X05000189

[18] L. L. Dorian Peters and N. Ahmadpour, "Toolkits, cards and games – a review of analogue tools for collaborative ideation," *CoDesign*, vol. 17, no. 4, pp. 410–434, 2021.

[19] R. J. Sage, *COGS: Communication Opportunity Group Scheme. Video*. University of Leicester, School of Education, 2000.

[20] A. Craft, "An analysis of research and literature on creativity in education," *Qualifications and Curriculum Authority*, vol. 51, no. 2, pp. 1–37, 2001.

[21] D. Van der Schyff, A. Schiavio, and D. J. Elliott, *Musical Bodies, Musical Minds: Enactive Cognitive Science and the Meaning of Human Musicality*. MIT Press, 2022.

[22] J. Freach, "Behold and beware, design toolkits," https://designcreativetech.utexas.edu/behold-and-beware-design-toolkits, accessed Nov 27, 2023.

IMMANENCE: ENVISIONING MUSIC IN A POST-ANTHROPOCENTRIC WORLD THROUGH VR REPRESENTATION

Chloë Sobek
Monash University
Melbourne, Australia
chloe@chloesobek.com

ABSTRACT

Immanence [1] is a work created for Renaissance violone, field recordings and electronics. This paper outlines the process of creating the digital animated notation, virtual reality (VR) score for the work. The score also functions as an immersive environment in which the audience experiences the audio-visual work via a VR headset. Additionally, I argue that a VR score that functions as an immersive experience has the potential to reinforce the broader conceptual concerns of a musical work. Moreover, it could further suggest a necessary expansion of the parameters by which be we define what music is and how it is represented, leading to new understandings and reflections on VR and musical composition. These ideas are underpinned by a response to Australian musician Jim Denley's provocations on contemporary music making, an ongoing commitment to considerations of post-anthropocentrism and a need to question predominantly humanist tenets that are still embedded across Western art music; the hope being that this paper could provide new methodologies for creating post-anthropocentric musical works.

1. INTRODUCTION

The motivation to create the VR score for this work was to provide a framework and provocation for my own compositional process. The idea was to create a VR score that represented sonic and conceptual parameters that I would like to see encapsulated in a musical work, but also to discover what unexpected, unplanned outcomes might arise from a score that was created in this way. In addition to this, I wanted to explore the possibilities of the VR score as an immersive environment in which to experience the work and whether this could help bolster the conceptual concerns of the piece; by which I mean broader philosophical ideas, beyond the scope of the sonic.

Having an interest in post-humanist and post-anthropocentric theory, VR as a representative medium additionally appealed to me because of its paradoxical nature. In her talk *Bubble Vision* (2018) [2] Hito Steyerl playfully points out the paradox of VR, that the viewer/audience perspective is always at the center of the scene, yet at the same time one's body is missing.[1] I find this paradox fascinating; the centering of the human within the VR medium is unequivocally anthropocentric even if it is merely imitating the way we corporally experience the world around us. Yet, in VR, the user's actual body is not present on the screen, making it simultaneously a curiously non-anthropocentric environment. I was intrigued by the challenge of building on Steyerl's paradox, by creating a work that attempts to subvert the notion of the disembodied human as the central most important figure, situated firmly at the center of every VR universe. I later discuss this idea in reference to the material remnants or artifacts of bodily gestures in the brushstrokes of paintings and the recordings of my own instrumental playing. Therefore, an additional motivation for this project is to create a musical work, aided by the medium of VR, that is post-anthropocentric, the decentering of anthropocentrism. These aims are sharpened by the ideas of Jim Denley and others that focus on challenging humanist notions and anthropocentric assumptions embedded in Western culture.

In this paper I describe the process of creating the VR score for *Immanence* using the animation program Blender [3]. I outline the audio-visual starting points and structure, the process of creating Blender assets, their animation and rendering the score for VR. I also discuss how I interpret the score in performance; as well as the various processing, sound spatialization and audio montage techniques I have employed. Additionally, I outline my efforts to represent the broader conceptual concerns of the composition sonically, and how the immersive environment reinforces this element.

This VR work was created within the confines of a few specific limitations, which I shall briefly outline now. Firstly, *Immanence* was created in 2D and was rendered first in a traditional video format. The issue this presents is that the work has been conceived in a linear form, within a specific aspect ratio. Future iterations of the work in VR would not be limited by these formalistic parameters. Secondly, *Immanence* has also been rendered in a basic VR version. I use the word 'basic' here as I have utilized the most rudimentary process for rendering a 2D work for VR

[1] Meta have tried to address this 'missing' element in VR by creating obligatory avatars. These avatars are odd to say the least but more importantly they still don't address the issue that one's physical body is never present in VR.

goggles. This paper highlights the current early, experimental stage of the work in development and incudes the navigation of these certain constraints, which means the work, for now, only provisionally functions as a 'score'. In future iterations, the VR score can be performed with live musicians and that the VR environment itself will be interactive for the audience, essentially providing the opportunity for the audience to 'play' parts of the score themselves.

I hope that this paper might contribute to the growing interest in employing the VR medium in the creation of musical scores and in art-making more broadly. I also hope to present a case for utilizing the VR score as the immersive environment in which to experience the musical work and that this might contribute to expanding the parameters of music representation and our definition of what music is.

2. BACKGROUND

I have a background as a double bassist and this instrumental practice has evolved into what is currently a composition and performance practice of enmeshed but disparate elements. The instrument at the center of this practice is the Renaissance precursor to the double bass, the violone. Drawing on musique concrète and improvisatory sensibilities I have become increasingly focused on the deconstruction and reconfiguration of instrumental performance using this instrument. This has evolved into experimenting extensively with augmenting the 'natural' acoustic sound of the instrument through electronic devices, granular processing, and extended technique. As an extension of this performance practice, I also create compositions through the process of slicing and manipulating samples of the instrument and collaging these to create audio-montages with other elements such as electronics and field recordings.

Musically *Immanence* does not depart from these approaches but rather utilizes the VR score to build on, expand and stretch these sonic ideas further and into new territories. Before I continue, it is worth mentioning that in no way am I attempting to argue that musical works need to be accompanied by a visual element. I believe quite the opposite and very much agree with the point Francisco Lopez makes when he insists that the audience wear blindfolds at his performances [4] and with Cat Hope when she states, "showing the score as a piece of music unfolds in performance detracts from the privileging of sound that a performance of music requires" [5]. This will always be an important stance when it comes to the presentation of musical works, and I will myself continue to make music that exists purely in the sonic domain. In this paper, however, I am exploring another, additional way in which to make and experience music, with the aim of stretching the term 'musical work' to encompass more than just sonic material, in this case an audio-visual experience. The act of labelling a work, that is experienced in VR, a musical composition privileges sound over vision as so much of the time it is the other way around, "as vision is almost never a mere accompaniment to the auditory" [6].

Another aspect to this work is the desire to subvert humanist notions that are often entrenched in western art music. The impetus for this came from Jim Denley's provocation:

> With human-induced ecological crisis, we face a crisis in music making — the narcissism we display in our self-referential temples, imagined engagement with place and anthropocentric music rituals, has lost all currency. Implicitly we understand we can't keep putting ourselves at the center (…) genuine new music, in our age, should provide alternatives to an anthropocentric world view. [7]

We are on the brink of existential catastrophe. Never has it been more important than it is now to question anthropocentrism and to think critically about the values that are reflected in our cultural works. Perhaps VR could provide an opportunity to create new 'temples' that are not the 'self-referential', 'narcissistic' spaces that Denley so deplores, but that are altogether different; spaces that are consciously constructed so as not to perpetuate humanist music rituals. In this work I attempt to represent these ideas sonically, aided, I believe by the medium of VR.

Immanence's success as a post-anthropocentric work also relies on the incorporation of field recordings, specifically those of the Australian bush and in a way follows on from Lindsay Vickery's response to Bailie's question, "if we are not to simply present the sounds of the world to an audience as a kind of musical fait accompli [...], what in fact are we to do with them?" [8]. Vickery proposes to "enter into a compositional and performative interaction with the recordings" [9], an idea I have interwoven into the composition of *Immanence.* I would argue that incorporating field recordings of the 'natural' world, the more-than-human world, into our musical works foregrounds the sounds and subjects of more-than-human others and thus contributes to the process of decentralizing ourselves and creating works that could be more post-anthropocentric in sensibility. Rosi Braidotti captures this idea in the following, "post-anthropocentrism displaces the notion of species hierarchy and of a single, common standard for 'Man' as the measure of all things. In the ontological gap thus opened, other species come galloping in" [10].

3. CREATING THE SCORE

The score for *Immanence* is both prescriptive and evocative. As Cat Hope points out, "composers engaging with animated notations seek something of a half-way place, where the most controlled and free can come together" [11]. This balancing act was central to the creation of this score and underpins the processes I describe in the following.

3.1 Paintings by Julian Aubrey Smith

As a starting point and path into the animated notation process, I worked with visual artist Julian Aubrey Smith on selecting works of his to base the score and digital animation on. I chose works that both encapsulated sonic ideas that I was interested in exploring and that presented new threads of enquiry to work with (Fig. 1). Furthermore, I was drawn to the painterly quality of the works. The brush strokes exhibited a 'handmade' element that were the material remnants of a gesture, and this felt like the perfect counterpoint to the digitally rendered world of the animated, VR score. Musically, I was interested in two ideas that were foregrounded for me in these paintings. The first being the concept of the 'portal' and the subsequent transformation; the transition from one place into another. The second aspect was the motif of the crumpled foil; a material form bearing the remnants of an action, as with the brushstrokes that were the materialization of a gesture. These were themes I wanted to explore musically and thus were elements that the score needed to feature and accentuate. Additionally, these material 'remnants' of bodily gestures (that I later echo on the violone) possess a kind of synchronicity with the missing body of the VR user; we know it is there or has been, but it is no longer visible nor a part of this particular VR universe. All that remains are its traces. I feel this element not only alludes to a type of decentering of the human, but also to a world where human beings are either no longer dominant or even present.

Figure 1. Julian Aubrey Smith. *Double Demi 1.* (2017)

3.2 Post-Anthropocentric Sensibilities and Representations

I have touched on some of the aspects I have included in *Immanence* that I believe assist in conveying the broader conceptual concerns of the composition but, aside from the use of VR and field recordings, there is another key aspect. The final component that I felt was necessary to include, was to create a sense of alienation for the audience; to construct an atmosphere where human ontologies weren't central to this VR universe. In other words, to create that feeling one has when stepping into the bush, the rainforest or any other environment not yet dominated by human beings, to replicate that feeling of curiosity, awe but also alienation that one might feel upon entering an environment such as this (Fig. 2). There are other VR/AR projects with similar aims, one such example is *Gardens of the Anthropocene* (2016) [12], which Lili Yan et al. describe as "appealing to the emotion of human vulnerability [...] presenting the erratic growth of mutated plants that outgrew buildings and fed on road signs beyond the control of humans" [13]. These concepts and requirements, combined with the other ideas mentioned previously, helped to guide the decisions when building the score in Blender.

Figure 2. Chloë Sobek. Still from *Immanence.* (2022)

3.3 Blender Assets, Animation and VR Rendering

The first assets I created were rectangular blocks to 'map' the photographs of Smith's paintings on to (Fig. 3). This is not the most straight forward process, as each image is added as a material and each material needs a system of nodes for it to be mapped with the right dimensions and coordinates to the object. I also created 'slabs of stone' that would represent the division of one space from the next.

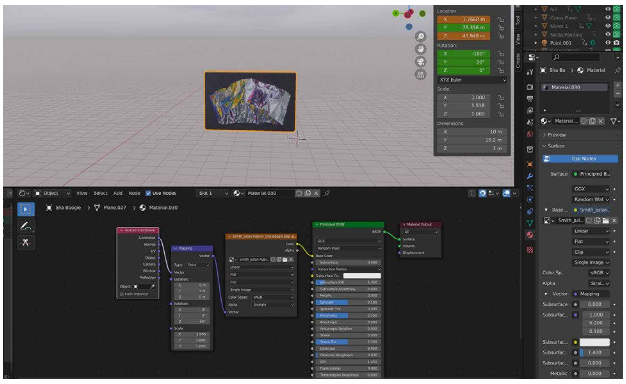


Figure 3. Chloë Sobek. Still from Blender Session*.* (2022)

Next, I worked on the assets, which were to become the animation environments or 'scenes'. Firstly, a grassy landmass (Fig. 4), which also involved nodes to map the different grass onto the plain. Then a sky and horizon through a plugin that could be imported into Blender. This brought a 'natural' looking light to the scene. Another environment consisted of a mirror like surface floating in the atmosphere, easily created by building a plain and adjusting its material properties to have a reflective surface. The final scene was again a landmass with foliage surrounded by a turquoise body of water, which was created from a plain that is again put through a series of nodes. The aim here was to create environments that were somewhat familiar but also strange; every element looks a little unreal and kind of kitsch.

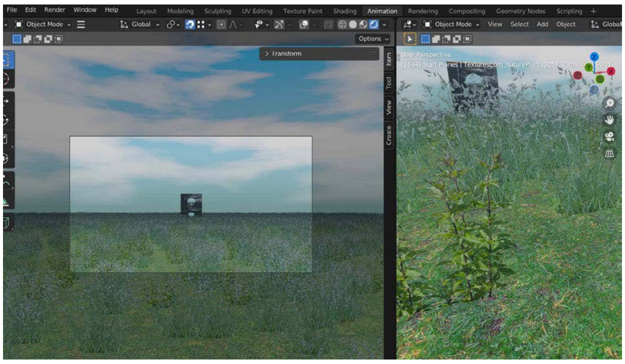


Figure 4. Chloë Sobek. Still from Blender Session. (2022)

The final Blender assets were objects that needed to exhibit seemingly autonomous behaviours. Again, to reinforce the idea that this VR universe possessed its own undisclosed system of operations and relations. I created a golden sphere, that reflected the grassy landmass around it, a series of crumpled, cardboard like cubes and a silvery, mottled asymmetrical object (Fig. 5) and (Fig. 6). These were simply adapted from the in-built Blender objects like the cube and spheres. I sculpted and stretched them and adjusted their material properties to be reflective or mottled, again through nodes. I also 'pulled' pieces of crumpled foil out from Smith's paintings (through photoshop) and used them as 'autonomous' objects as well.

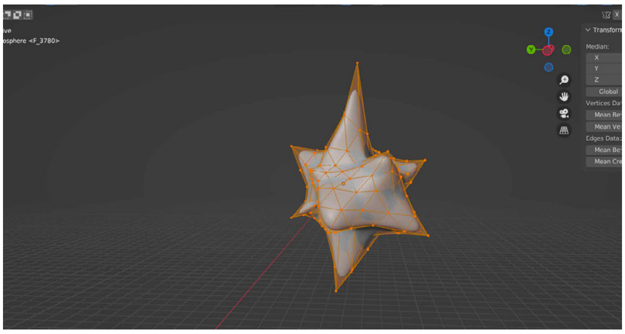


Figure 5. Chloë Sobek. Still from Blender Session. (2022)

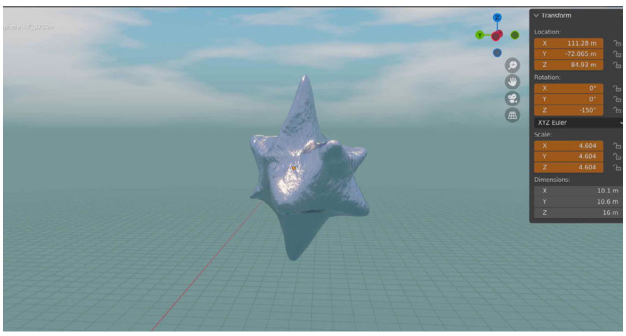


Figure 6. Chloë Sobek. Still from Blender Session. (2022)

With the Blender assets built I started the animation process. This aspect was perhaps the most important with regard to ensuring the Blender work functioned as a score. The animation process was slow and involved moving each object individually then inserting keyframes (Fig. 7), including the camera (aspect) and lights. Every decision in the animation process/score creation was made according to three distinct parameters. These were the concepts gleaned from Smith's paintings, the sonic elements I wanted included in the work and post-anthropocentric sensibilities and representations.

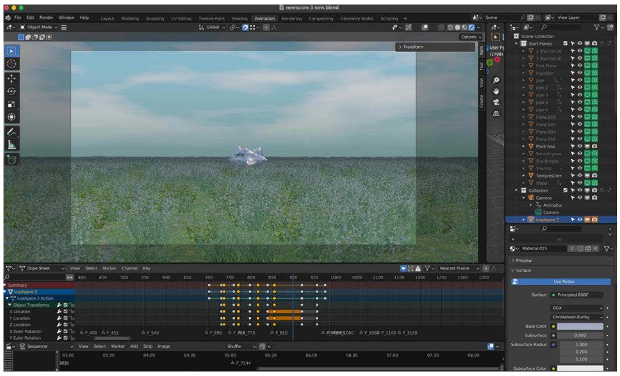


Figure 7. Chloë Sobek. Still from Blender Session. (2022)

Each aspect covered in the following provides the directives for the musical interpretation that will be discussed in the next section. The first and second painting in the score both depict what could be perceived as portals into other places (Fig. 8). In response to this, the animation starts in one environment with the first painting being gradually revealed from behind the 'slabs of stone'. As the painting is revealed the viewer moves closer and closer towards the 'portal', eventually travelling through to the other side. In the next painting we see a golden sphere mimic this movement by emerging from the portal and transitioning through the space towards the viewer.

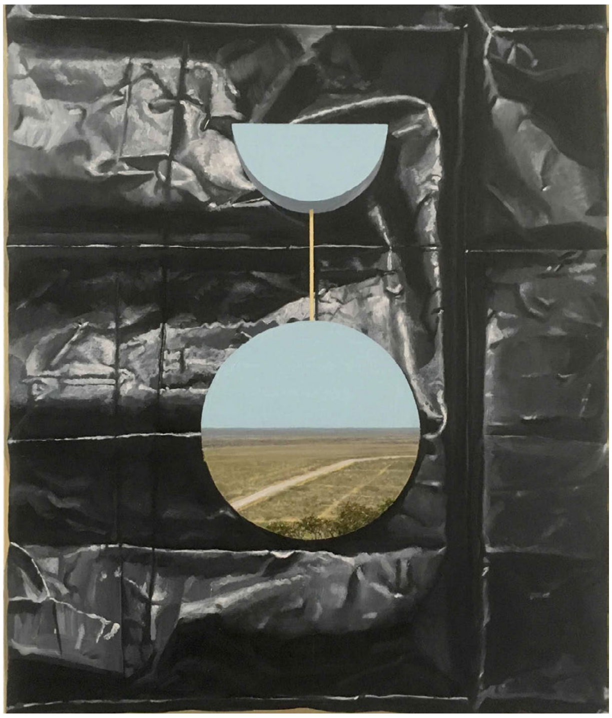


Figure 8. Julian Aubrey Smith. *Double Demi 2.* (2017)

The remaining paintings in the score feature more specifically the crumpled foil motif, which were highlighted through 'extracting' parts of the paintings and animating them. I also highlighted the brushstrokes in the paintings as, having mentioned previously, these were an important feature of the work. I emphasised these by bringing the camera aspect very close to the paintings to capture the brushstrokes in detail.

In addition to the concepts the paintings introduced to the score, I added audio-montage and sound spatialisation. Audio-montage was represented by the layering and existence of the different Blender assets in the scene. Sound-spatialisation was represented by the immersive environment. Here VR being a completely 360-degree, immersive environment was paramount. Sound spatialisation can in VR be represented spatially as it is heard, this is a unique aspect of a VR score and very much one I did not utilise enough in this work. I animated some of the objects so their behaviour in space could be followed sonically and I animated the camera view at points to encourage a shifting or swirling of the sound across the stereo field. This resulted in interesting sonic outcomes. I used some electronic sounds that were texturally dense, with a harshness that, coupled with the whirling camera view, is almost sickening.

The final task was to utilize remaining aspects of the animation to reflect post-anthropocentric sensibilities and representations. I did this by inserting unexpected aspects to the camera animation like 'whirling', or dropping, or floating in space. At one point being propelled through the foliage and eventually being lowered into the water. These elements were intended to create a sense of uncertainty, surprise and even lack of control for the audience.

With the animation complete, it was rendered for VR. This was done through Blender, then eventually though 'Spatial Media Metadata Injector'. I would not recommend this process as it took a lot of computer processing and trial and error. For future projects I would use a program like Unity or Unreal.

3.4 Designing the Directives of Musical Interpretation

As mentioned earlier the score involved the balancing of prescriptive elements and more open, evocative directives. The instructions were as follows:

- The overarching concept behind the score is to create a post-anthropocentric sensibility, in other words, to capture a somewhat alien environment, a world eerily absent of any perceivable human life. The score attempts to replicate a feeling of curiosity, awe, but also the alienation one might have when walking into the wilderness or any other environment that is not yet dominated by human beings. This overarching concept and aesthetic influenced the creation of the score and should thus influence the musical interpretation of the score.
- The visual aesthetics and evocative quality of the score should match the sonic interpretation.
- The content of the paintings needs to be responded to sonically.
- The work must include fielding recordings of more-than-human sounds.
- Each distinct place in the score needs to be accompanied by different sonic material.
- The transition between spaces needs to be represented sonically through the merging of material or ending it abruptly.
- The movement of objects and the camera aspect can be accented or accentuated by the spatialisation of the sonic material at will.
- Brushstrokes and crumpled foil are akin to sounds being physically created on an acoustic instrument and therefore there should be an instrumental element throughout the piece.

4. INTERPRETING THE SCORE

The interpretation of the score did not unfold in the traditional sense of being performed in one live take but instead relied on various recording practices and musical collage techniques and therefore had to be planned out to a degree. I recorded my own interpretation and presented it as an audio-visual experience in a concert environment. I did this so that the work could be experienced succinctly; all elements contained within the VR headset. In future iterations it would be interesting to perform the score live while the audience experienced it in VR.

Observing the score instructions, I interpreted the score in the following ways, dividing the approach into three categories again. The paintings and accompanying animation, the specifically musical elements, and the post-anthropocentric, evocative elements. There are four distinct 'spaces' in the score and each of these were divided by either symbolic 'stone walls' or a transition represented by a sort of 'falling' through space (transitioning through 'portals'). Musically each space was distinguished by unique sound worlds that were created by crafting the sections separately then collaging them together or ending a sonic segment abruptly. Notions of transformation and 'the portal' were represented by sounds that, as mentioned, either morphed and shifted at their transition point, bleeding into each other but remaining distinctly different or ending abruptly with a recurring drum motif. The recurring crumpled foil and brushstrokes in Smith's work (the remnants of a gesture or action) were represented in the violone parts. Though temporal objects, the violone sections mirror this 'remnants of action' idea in the sense that the sounds cannot be disentangled from the bodily gesture that created them. As mentioned earlier I also incorporated elements of sound spatialisation that at times followed the movement of the objects or that highlighted that 360-degree environment with the sound being spatialised around the audience.

Finally, it should be mentioned that my attempt to sonically represent post-anthropocentric sensibilities (just as was the process of representing these sensibilities visually) is merely a starting point and these approaches are experimental at best. Here I outline the ways I went about doing this. The score depicts a simultaneously familiar and unfamiliar space, with intangible logic, slabs of stone, strange portals, autonomous objects. There are landmasses of luscious foliage and turquoise waters with strange tentacular objects that rise to the surface. These informed the musical interpretation; a field recording of fruit bats, an explosion of seemingly symbolically significant drums, heralding something unknown, undisclosed. A feathery bass drone accompanies the eerie sounds of violone harmonics and horsehair against the wood of the instrument, which morph into the somewhat alarming sounds of a Kookaburra. It is worth mentioning that the field-recordings are unmistakably Australian sounds, more specifically, sounds from Yarra Bend Park and the Mornington Peninsular in Victoria. This means the field-recordings are very much situated in these distinct places. This is in stark contrast to the very unspecific, dislocated, even generic visual VR world that is the score, further emphasizing that sense of discomfort and dislocation for the audience. The work also opens with the drum motif, designed to startle, and create a sense of discomfort from the beginning of the piece. This is followed by the disembodied sound of fruit bats, spatialised as if they are swarming and whirling around behind the viewer, yet never visible. Towards the end of the work, in response to the animation dragging the viewer through foliage and eventually descending into the water, I included a section in which I had layered several violone improvisations over one another. I processed this with a type of granular synthesis. The effect is that of something broken, like a recurring but inconsistent glitch, adding to the chaos and discomfort as the work drew to a close and the audience are gradually submerged in a body of, now swamp green water.

5. CONCLUSIONS

Immanence has demonstrated that a VR score has the potential to provide new and interesting representational structures for creating musical compositions and that these can be both prescriptive and more open and evocative. The use of VR in this project has also facilitated new ways to incorporate visual work such as paintings into a musical score and respond to them in an immersive way. A VR score can be a richly generative immersive space in which to score sound spatialization, especially for headphones since the score can visualize the placement of the sound in a 360-degree environment. I have also argued that the VR score can simultaneously act as an immersive environment in which the audience can experience, and potentially interact with the work. This could contribute to the development of future musical rituals, spaces and 'temples' that have the potential to call in to question existing, possibly outdated models. *Immanence* has also demonstrated how utilizing the VR score additionally as an immersive environment in which to view the piece can assist in communicating the broader conceptual concerns of a musical work. Cat Hope defines animated graphic notation as "a predominantly graphic music notation that engages the dynamic characteristics of screen media" [11]. Perhaps we could expand this definition and state that animated graphic notation is also something that engages the evocative potential of 360-degree environments and can even provide a new space or 'temple' in which to experience musical works. My definition of animated graphic notation therefore is: 'Animated graphic notation is a form of graphic, musical notation that employs the kinetic elements of screen media as a set of directives, but can also be the very 360-degree domain in which to experience the musical work itself'.

One of the limitations of this project is that it is yet another musical work with a substantial visual component; the danger being that this could potentially distract from the sonic element, which is intended as the focus of the work. However, VR is a digital technology that is becoming ubiquitous and is, therefore, a medium worth experimentation and exploration, even for artists working with sound. Mark Zuckerberg states, "there are 200 million people who get new PCs every year, primarily for work. I do think that [...] you're going to be able to do pretty much everything and more you can do on PCs on VR" [15]. As Cat Hope states, "animated notation leverages the contemporary digital technologies that are already a large part of our lives" [11]. As Meta aggressively pushes their VR agenda, we need ways to contemplate and conceptualise both the positive and negative aspects of this technology. Exploring the potential of creating musical scores in VR and experiencing musical works in these types of immersive environments is certainly a starting point in this process.

Acknowledgment

I would like to thank Cat Hope and Iran Sanadzadeh for their ongoing support and DigiScore, who generously provided the opportunity for me to create this VR score. Thank you also to Hamburg Institute for Advanced Study, European Research Council and Monash University. Finally thank you to Julian Aubrey Smith and Vijay Thillaimuthu.

6. REFERENCES

[1] C. Sobek, "Immanence," (2022). [Video]. Available: *Immanence Non VR Version.mp4* *Immanence VR Compatible.mp4*

[2] H. Steyerl, "Bubble Vision," *YouTube,* Jan 31, 2018. [Video file]. Available: https://www.youtube.com/watch?v=T1Qhy0_PCjs

[3] The Blender Foundation, "Blender," [Online] Available: https://www.blender.org

[4] F. Lopez, "Interview for DB magazine," June 2004. [Online] Available: http://www.franciscolopez.net/int_db.html

[5] C. Hope, "Electronic Scores for Music," *Computer Music Journal,* vol.41, no. 3, Fall 2017, pp. 21–35.

[6] C. Cox, *Sonic Flux: Sound, Art and Metaphysics.* University of Chicago Press, 2018.

[7] J. Denley, "Hollis Taylor's Absolute Bird 'Moving beyond being quite obsessed with ourselves' ", [Online] Available: https://www.australianmusiccentre.com.au/article/hollis-taylor-s-em-absolute-bird-em

[8] J. Bailie, "Squeezing out the music from real sound". Interpretations. Q-O2 Winter, 2014. [Online] Available: http://www.q-o2.be/uploads/interpretations.pdf

[9] L.Vickery, "Notational Strategies for Integrating Live Performers with Complex Sounds and Environments," in *Proceedings of the International Conference on New Tools for Music Notation and Representation (TENOR '20/21),* Hamburg, 2020/21.

[10] R. Braidotti, *The Posthuman,* Cambridge, UK: Polity Press, 2013.

[11] C. Hope, "The Future is Graphic: Animated notation-for contemporary practice," *Organised Sound*, vol. 25, no.2, 2020, pp. 187–197.

[12] T. Thiel, *"*Gardens of the Anthropocene*,"* (2016).

[13] L.Yan, M. Colleni and B. Litts, "Exploring the Rhetorical Affordances of Augmented Reality in the Context of the Anthropocene," in *Proceedings of 6th International Conference of the Immersive Learning Research Network (iLRN 2020),* Online June 21–25, 2020.

[14] A. Heath, N Patel, "Mark Zuckerberg on the Quest Pro, building the metaverse, and more," Oct 2022. [Online] Available: https://www.theverge.com/23397187/mark-zuckerberg-quest-pro-metaverse-interview-decoder.

SKETCH, DRAFT AND REFINED HYPOTHESIS TESTING AS A CREATIVE PROCESS IN AUDIO PORTRAITURE

Maree Sheehan
Te Wānanga o Aotearoa, New Zealand

ABSTRACT

Audio-portraiture describes a discrete, immersively experienced, audio rendering of an individual's identity. Paradigmatically the development and creation of these audio portraits is informed by te ao Māori philosophies, knowledge and understandings. Within a process of reflective practice, I can trace the iterative development of my portraiture through three forms of experiments, audio-sketching, audio-drafting and refined hypothesis testing. These three forms of experimentation offer a progressive artistic practice-led approach of a non-linear fashion.

Immersive sound technologies are utilized to record, synthesize and spatially position interpretations of the person's perspectives, experiences and nature. This paper provides an overview of how audio-portrait provide an artistic synthesis of sonic practices resulting in activating sensory responses for a listener that reach beyond the parameters of visual portraiture. This is because 360 immersive and binaural sound-capture technologies can be orchestrated into artistic works that convey unique experiences of space and time. Such work may be designed as a distinctive form of portraiture. Significantly, audio portraiture demonstrates the multidimensionality of wāhine Māori *(Māori women)* through artistic interpretation and representation in ways that are culturally authentic and unique.

1. INTRODUCTION

The interpretation and representation of the multidimensionality of wāhine Māori *(Māori women)* through audio-portraiture posits through artistic practice, an approach where one might integrate the physically accountable (identity, knowledge, recollection, opinion, and music) and the esoteric. As such, it positions wairua[1] *(spirit)* and mauri[2] *(life-force)* as living, communicable phenomena, capable of interpretation. It suggests that the immersive nature of sound has the potential to activate sensory responses for a listener that reach beyond the parameters of visual. This is because 360 immersive and binaural sound-capture technologies can be orchestrated into artistic works that convey unique experiences of space and time [4]. Such work may be designed as a distinctive form of portraiture. This constitutes a distinctive renegotiation of how wāhine Māori might be interpreted and, in so doing, they disrupt a largely visual concept of portraiture that was imported into Aotearoa/New Zealand during the process of colonization [5].

My creative practice takes the form of three audio-portraits that have been the result of iterative experimentation and reflection. As an artistic practice it, "is concerned with the nature of practice and leads to new knowledge that has operational significance for that practice" [6, p. 1]. The portrait is composed entirely of sonic elements. Thus, sound becomes the artist's palette with the sound design becoming the compositional experience and canvas. A listener's experience is such that they hear, feel and sense the essence of the wāhine in an audio-portrait that is composed within a 360-sound environment of simultaneous sound cues [5].

Paradigmatically the development and creation of these audio portraits are informed by a Kaupapa Māori methodology including te ao Māori philosophies and our ways of expressing thought and being, including waiata (*song*), karakia *(incantation, prayer)*, mihimihi *(to greet, pay tribute)* and whakataukī *(proverbs)*. Kaupapa Māori as a philosophy that guides Māori research [7, 8, 9, 10]. Such a paradigm locates Māori understandings and philosophical beliefs as central to processes, analyses and intended outcomes. A Māori worldview encompasses connections to, and understandings of all things Māori, from histories, traditions, spiritual understandings, tikanga *(Māori customs and practices)* and the use of te reo Māori *(Māori language)* to transmit cultural knowledge. Employing a Kaupapa Māori paradigm has provided a useful research orientation that utilizes familiar cultural values. These values have proven effective because I am a Māori woman working with other Māori women, but at the same time, my research draws upon the knowledge of non-Māori contributors.

Co-jointly, this project may also be conceived of as an artistic inquiry. I refer to a process of iterative development of thinking through immersion within conceptual and material development. It is artistic practice that 'leads' the research through its ability to raise questions and discoveries that shape refined versions of the audio portraits. The

[1] Mead translates wairua as 'soul' or 'spirit' [1, p. 59]. Reverend Māori Marsden's definition, where he suggests, "wairua (spirit) or hau (the breath of the divine spirit) is the source of existent being and life." [2, p. 47].

[2] Marsden defines mauri as a "life-force" [2, p. 95]. Mauri in reference to the force or energy that holds or bonds all things together and is the connection to everything [2, 3]. Mauri may be understood as the life principle of an object, individual or ecosystem and its essence or force [2].

concept of reflective, practice-led research relates to Donald Schön's concept of reflective practice [11]. According to Niedderer & Roworth-Stokes "emphasises the role of reflection within a process to derive new insights and understanding to further research and professional practice" [12, p. 10]. An approach of reflective practice provided a framework that was integral for both iterative development and diverse forms of thinking and rethinking. It supports the time to dwell prior to beginning the creative process (and at times during it), which reinforced connections to my creative self and validated time to imagine, reflect and be in a state of openness to both the physical and non-physical and that connection to the three wāhine Māori whose audio portraits I was developing and creating [5]. Drawing upon these processes of reflective practice, I can trace the iterative development of my audio portraiture through three forms of experiment: Audio sketching, audio drafting and refined hypothesis testing.

2. AUDIO PORTRAITURE

Davidson, Seaton and Simpson describe a portrait as a rendering of "the likeness of a real person: a vivid description" [12, p. 768]. Their term 'vivid' suggests something beyond a technically figurative reflection. It proposes that a portrait might reach beyond the surface into something richer and more profound. Significantly, their definition does not assume that a portrait is a visual document. Portraiture been discussed in diverse mediums including painting [14, 15], photography [16, 17, 18, 19], narrative [20, 21, 22], and multi-media [23, 24, 25].

Building on these ideas of diverse portraiture interpretations, I define an audio portrait as an original rendering of an individual's identity. Such portraiture seeks to respond to the physical and the spiritual nature of the wairua and mauri of particular wāhine, through the creation of a sonic immersive experience that utilises binaural and immersive sound technologies. Audio portraiture contributes to the conceptualization and exploration of a unique, emerging media form that expands on the concept of portraiture as an aural medium [5].

The rationale of developing and creating audio portraits of wāhine Māori comes from the need to address a dominant, colonially constructed and under questioned, mode of representing Māori women. Drawing on Māori worldviews, this study rethinks how we might perceive the complexity of women as both physical and non-physical beings. readdressing the effects of colonisation and the imposition of non-Māori concepts of race, gender and class. These dramatically changed traditional Māori belief systems in relation to wāhine. Colonial belief perspectives positioned Māori women as inferior to Māori men [26, 27]. The negative effects of colonisation resulted in misguided assumptions that resulted in both Māori women and atua wāhine *(goddesses)* being misinterpreted, censored or made completely invisible.

Importantly, portraiture of wāhine has remained largely confined to the concerns with pictorial imagery and as such, it has failed to draw into consideration the potentials of a rich spectrum of purely aural modes (including kōrero[3], karanga[4] and oro[5]) that are integral to Māori ways of knowing and being. This is a significant issue when we consider that identity depiction of wāhine has a long and rich history in oratory, waiata *(song, chant)* and pūrākau *(narrative)* and taonga puoro *(traditional Māori instrumentation)*.

Audio portraiture grounded in Māori ways of knowing, explores specifically how audio-portraiture might capture and embody the *essence* of Māori women through sound, by responding to multiple dimensions of their identity. The honouring and affirmation of of wāhine Māori voices and the assertion of Mana wahine[6] explores how audio-portraiture might provide a way of reconceptualizing biographical material within a Māori epistemological framework by integrating the physically accountable and the esoteric.

2.1 Audio Depictions of Identity

From this corpus of research, specific ideas have proven useful to my inquiry. Miller's [30], Friðriksdóttir's [31] and Ndikung's [32] discussions of sonic portraits as compositional artefacts that draw on diverse repositories of sound, including the vernacular and intimate, resonate with my audio-portraits. This is because like them, I integrate diverse material from a broad spectrum of sonic data, including published music, laughter, intimate conversation and the ambience of a marae *(grounds, land that belong to a particular tribe or sub-tribe),* street or coastline. Although I am not concerned like Gridley [33], with interfaces between visual portraiture and sound, like her, I understand that sound has a significant relationship with the nature of the participants with whom I am working.

At times music is an integrated component in my portraiture, unlike Bach, Sadja [34] and Bongiovanni [35], I do not conceive my work as musical portraiture. But, like Bongiovanni and Gridley, I seek to explore, express and represent the 'essence' of the women I am interpreting, as such, like Walden's description of Bach [36], my audio-portraiture is an attempt to represent an individual's character, rather than her physical likeness.

I see my work as composed sound that might be positioned within a broader conception of Sound Art [37]. Thus, I see myself as a sound designer, audio culturist and sound practitioner whose audio-portraits constitute creative audio expressions of material found in the natural world, integrated into sound designs and physical and non-physical ecologies [38]. Here my audio-portraits function as a social creation existing inside a complex social network [39] and these works function as a form of "sonic resistance" against colonial practices [32]. In composite, the audio-portraits contribute to Cox and Warner's 'new sonic landscape' where they may be seen as revealing a rich ontology of sound [40].

[3] Kōrero – "to speak" [28, p. 136].
[4] Karanga – "to call or summon" [29, p. 98].
[5] Oro – "sound" [29, p. 242].

[6] "Mana Wahine theory is a Kaupapa Māori theory that is dedicated to the affirmation of Māori women within Māori society, within whānau, hapū and iwi" [9, p. ix].

2.2 Sound Technologies

Within the development of audio sketches, audio drafts and refining hypothesis testing, immersive and binaural sound technologies have provided a system to create an embodied sensory sonic experience of an audio portrait. By employing binaural and immersive sound technologies in the design of the audio-portraits, I was able to create an immersive spatially sonic experience. The interpretation of each wahine within this space emerged through techniques that simulated intimate hearing cues. These binaural sound technologies included mono, stereo, binaural and ambeo audio data capture and the utilization of Logic Pro X and Pro Tools HD with Facebook 360 (for the arrangement and mixing of specific features).

The audio portraits are constructed in a sonic 360 space. This space may be understood as immersive. By this I mean when experiencing these portraits, one moves beyond the experience of stereo listening to a state where sound moves around and through the body. The term immersive sound relates to spatial sound quality that produces a lifelike sense of being immersed in the presence of people or environments. Sound is localized at various distances from the listener [41]. This orchestration requires audio material to be positioned within a 360-surround sound sphere environment,[7] which can enable us to listen '*within*' rather than '*to*' sound, such that we experience a sense of sound immersion. When we are immersed in sound, spatial distances may be experienced above, below, adjacent to, or moving through the listener's body [5].

Roginska and Geluso [42] describe binaural sound as two-channel sound that enters a listener's left and right ears at the same time.[8] Such sound has been filtered by a combination of time, intensity and spectral cues and is intended to mimic human localization perceptions. As Møller [43] notes, "The input to the hearing consists of two signals: sound pressures at each of the eardrums. If these are recorded in the ears of a listener and reproduced exactly as they were, then the complete auditive experience is assumed to be reproduced correctly" (p. 171). This type of audio recording attempts to "accurately copy the way in which humans perceive sound, ensuring that sound waves reaching the head undergo the same transmission on their way to the ear canals" (p. 172). Binaural sound is generally reproduced and experienced through headphones but can also be simulated through speaker systems.

2.2.1. Field recording and data gathering

A variety of recording devices for capturing ambient, ecological and atmospheric data, alongside human vocalizations, such as interviews, singing, humming and breathing were utilized. Sound data also included capturing the sounds of home life, the sounds of cultural significant environments such as the sounds around marae *(ancestral tribal meeting grounds)* which included awa *(river or lake),* wind, trees, insects, birds and other ambient life forms. The sound recording equipment included the Sennheiser ambeo VR microphone, the Neumann KU100 dummyhead, the Sennheiser MKE600 directional microphone and the Aquarian H1a hydrophone. The Zoom F8 field recorder provided 8 inputs which provided a way to record all these microphones at once when necessary.

3. REFLECTIVE PRACTICE

Reflective practice as a research methodology was developed by Donald Schön between 1983 and 1987, and in the development and creation of the audio portraits, it functions as a means to refine thinking inside a process of iterative experimentation. The first Schön calls an "exploratory experiment" [11, pp. 141–149]. This is when an action or series of actions is taken with the purpose of seeing what emerges. This form of experiment tends to occur after data gathering (in my cases kanohi ki te kanohi (*face to face)* and ambient environmental recordings) when initial 'audio sketches' are generated as a way of exploring potentials within a broad understanding and interpretation of the subject of the inquiry.

The second form of experiment Schön defines as "move-testing" [11, pp. 146–147], which involve an action that is undertaken in a deliberate way to make a specific change. Such moves are considered in terms of the whole and after critical reflection; they are either negated or developed into a more sophisticated text. This form of testing or 'audio-drafting' was used while I was editing the audio-portraits, and in the development of spatial immersive nuances of sound within the design.

Schön also posits a third type of experiment that he describes as "hypothesis-testing" [11, pp. 143–144]. Scrivener [44, p. 6] suggests that such an experiment, "succeeds when it reflects an intended discrimination among competing hypotheses." This type of experiment is evidenced in instances where several possible versions of the same audio-portrait are refined. These potential approaches are then analyzed, compared and refined in pursuit of the most effective interpretation of the subject's identity.

Within a process of reflective practice, I can trace the iterative development of my portraiture through three forms of experiment:

- Audio-sketching (where the researcher tests in a series of audio 'thumbnail sketches' to see what might be discovered) [Schön's exploratory testing]
- Audio-drafting (where thinking within a work is iteratively developed and trialed) [Schön's move testing]
- Refined hypothesis testing (where advanced experiments are compared and evaluated before one is rendered as a completed portrait) [Schön's hypothesis testing].

Although these three forms of experimentation may appear to offer a seemingly objective, delineated process, this

[7] This sound sphere environment is created using Pro Tools HD and Facebook360 Spatial.

[8] The binaural microphones used to capture such recordings are referred to as a dummy head.

is often not the case. I do not always progress thinking in a purely linear fashion. For example, it is quite common for me to respond to reflection on a hypothesis or audio-draft by returning to a process of audio-sketching to explore and broaden the range of considerations or audio within a portrait.

The concept of reflective practice is not new. In addition to Schön; Boud, Keogh and Walker [45] observed that reflection inside and through practice, can be part of the way an individual uses experiences to lead into to new conceptual perspectives or understandings. McClure [46, p. 3) argues that "reflection is a complex concept that has defied consensus of definition although some commonalities exist". These shared features she suggests are the centrality of the self whose problem-solving abilities are triggered by questioning "actions, values and beliefs".

Inside my research processes I engage with two distinct forms of reflection: 'Reflection *in* action' and 'Reflection *on* action'. Reflection *in* action occurs whilst working *inside* a problem that is being addressed, (in what Schön calls the 'action-present'). Høyrup and Elkjaer see such reflection as "a discursive way of creating a space for focusing on problematic situations and of holding them for consideration without premature rush to judgment" [46, p. 23]. In such a state, I reflect on thinking as it emerges through the process of embodied practice. Here, I often draw on tacit knowing and thinking that may not be verbalized.

Conversely, reflection *on* action occurs *after* an experiment, when decisions have become manifest. Such reflection is consciously and critically undertaken, it is evaluative and I often use it as a process for gaining a more dispassionate and articulable overview of my decisions.

In both forms of reflection, I engage in a process similar to that described by Ings [48, p. 80] that involves a

> complex level of analysis and synthesis within emerging bodies of data. Within this, [I] continuously investigate potential patterns, parallels and associations within the information, and from this, [I] project new questions back into the project.

This kind of processing and analysis Ings argues is, "subjectively reflective and transactional" and this subjectivity is intimately connected to my cultural values, feelings and experiences. I hear sound, then I think and 'feel' how a portrait might be developed. Conscious and unconscious processes operate co-exist inside this transactional environment, where I communicate with the emerging data and it talks back to me. Inside this process, I draw together what is imagined with what can become evidenced. Scrivener [44, p. 7] in describing this process says:

> there is a recognition that the creator's interest is in trans-forming the situation (i.e. psychological, emotional and created) to something better (e.g., equilibrium between intention and realisation).

4. CREATIVE PROCESS, SYNTHESIS, IDEATION AND FLOW

The development and creation of audio portraits involves the creative synthesis and processing of audio recorded data. Data synthesis utilizes three broad methods; immersion, iterative development and critical review. Because immersion and flow permeate the entire creative process, it is useful before discussing the actual 'methods' employed in the iterative development of the portraits, to consider these two states of being and thinking.

4.1 Immersion

Immersion refers to a process of indwelling or entering an inquiry in a manner where the question and environment are internalized [49, 50, 51]. Once data is collected, I draw it into myself and contemplate its potential. In this process I dwell in a spiritual dimension that supports my connection to wairua and mauri, as well as my internalized creative process. Pere describes this as "a dimension internalized within a person from conception – should the seed of human life emanated from Io, the supreme supernatural influence" [52, pp. 13–14].

According to Douglass and Moustakas [49] argue immersion allows for an intensive and comprehensive understanding of a particular moment or experience. In this sense, they describe immersion as a particular state of mind where the researcher becomes integrated with the research problem in an auto-centric mode, where aspects of the researcher's 'being' are centered on the theme investigated. They suggest that in this state "vague and formless wanderings are characteristic in the beginning, but a growing sense of meaning and direction emerges as the perceptions and understandings of the researcher grow and the parameters of the problem are recognised" (p. 47). They also argue that, for immersion to operate productively, the researcher "must stay in touch with the innumerable perceptions and awareness that are purely [her] own".

4.2 Creative Process-Flow

Although a process of immersion enables me to draw correlations between data and the self, the act of creative synthesis is distinguished by a pronounced sense of flow that constitutes a forward momentum in creative thinking. Csikszentmihalyi [53, 54] has discussed such 'flow' as:

> our experience of optimal fulfilment and engagement. Flow, whether in creative arts, athletic competition, engaging work, or spiritual practice, is a deep and uniquely human motivation to excel, exceed, and triumph over limitation. [54, p. 266]

He sees creative flow as a unified and coordinated approach where "attention becomes completely absorbed into the stimulus field defined by the activity" (p. 239). This state of complete involvement permeates my artistic process. It is, "an intense experiential intricacy, weaving moment by moment the focus and attention to the person's fullest capacity so as to create (p. 239). I experience flow as spontaneous and effortless, yet paradoxically it requires an immense amount of concentration. When in a flow, I am completely engaged with artistic creation, everything else becomes peripheral. In this state, the world of the emerging portrait becomes my world; I possess and am possessed by the process of creating, experimenting with, and evaluating something that is emerging from the undefined into the defined. In maintaining this autotelic state[9] I am careful to isolate myself from distraction, I lose a sense of time and the deeper I am immersed, the more momentum I achieve.

5. SKETCHING, DRAFTING AND REFINING

From a state of immersion, moving outward into a process of flow, the iterative development of my audio-portraiture passes through three forms of experimentation: audio-sketching, audio-drafting and refined hypothesis testing.

5.1 Audio Sketching

Audio-sketching is a form of initial experimentation where I create thumbnail sonic sketches using audio recordings and archived sound. So, an audio-sketch may be likened to a thumbnail sketch or a rapid assembly to see what I might find. In this exploratory approach, I listen to the kōrero of each wahine and I look for what expresses her understandings of identity. I also try to discern what is important to her socially and politically. Here, I am attentive to nuances of her mauri and wairua, as they surface through her kōrero, her breath, her humming, her singing, her laughter and the environmental and ambient sounds that surround her. I also draw on music and archival material that she has composed. These are interwoven in rapid assemblies that are initial responses to the question 'What is the essence of this woman?'.

When developing audio-sketches I will sometimes explore the potential of taonga puoro *(traditional Māori instrumentation)* such as the pūtōrino and the kōauau, as a way of complementing or 'speaking with' the each wahine, her whakaaro and kōrero. Taonga puoro can also reflect a connection to her whenua *(land).*

Throughout the audio sketching process, I trial synthesized samples sounds from the Logic Pro X sound bank, or other media. I may also embed reflective expressions of what I have experienced in small compositions of my own guitar, piano or synthesizer and sampler playing.

Audio-sketches are arranged and rearranged in a process of 'listening-experimenting-listening' to produce quick compilations that test the potential of diverse sound arrangements and emphasizes. These experiments are normally less than two minutes in duration and are utilized as reflections, although some may expand into audio-drafts, which are more substantial sound arrangements. After listening to an audio sketch, I reflect on what it sounds and feels like. At this point, I either choose to preserve or delete a composition or retain parts of it [5].

Please see example of audio sketch of Moana Maniapoto and the end of this paper.

5.2 Audio Drafting

An audio-draft, as distinct from an audio sketch, may be compared to the manner in which a drawing or 'study' may be differentiated from a rapid pencil sketch. An audio-draft is essentially a deeper orchestration of elements that involves a critically expanded and orchestrated palette of audio material.[10] In constructing an audio-draft an extended amount of time is spent designing and critiquing. The experiment therefore engages a more sophisticated level of syntheses and refinement. Before embarking on an audio-draft, I immersed myself in initial sketches, listening to the nuances of voice, narrative and musical soundscapes. I did this to consider how combinations of sound have conveyed certain emotions or captured an expression of mauri or wairua. I then rearranged these sounds; adding, subtracting, elongating or shortening elements. Although I continued to use Logic Pro X, I introduced sound effects or plugins to experiment with reverb, delay (echo) and equalization (EQ). I also utilized Pro Tools HD for experimental mixing in a 3D audio workflow. When audio-drafting, I also experimented with the positioning of various sounds within 360 sound spheres. I listened back on headphones to hear how these locations of sound might create an immersive, evocative audio representation of the wahine. These 360 immersive sound spheres are created and reworked through a process of move testing, until I think, feel and hear that I have created a resonant interpretation [5].

With audio-drafts I am pursuing the explicit and esoteric nature of each wahine. As an audio-portrait begins to evolve, it develops a mauri of its own. Through deep listening, sound arrangement, sound creation, editing and mixing processes, I responded to this mauri … trying to acoustically draw it into an audio form [5]. On a non-esoteric level, this process may be likened to Schön's subjective transactional relationship [11] because the audio-portrait talks to me and in so doing, it is shaped and developed into more refined form. Here, I am in deep and sustained conversation with the portrait.

Please see example of audio draft of Moana Maniapoto and the end of this paper.

5.3 Hypothesis Testing: Finalizing Audio Portraits

The final form of experimentation is similar to Schön's 'Hypothesis testing'. This type of testing is used to "effect an intended discrimination among competing hypotheses" [11, p. 146]. In my creative process, it is used to compare the qualities and potentials of more than one refined version of the portrait of each wahine.

Hypothesis testing experiments are considerably more technically advanced than audio-drafts. Developing hypotheses of audio-portraits requires sound mixing in Pro Tools | HD in a 3D spatial audio work-flow environment

[9] Self-resourcing and having an end or purpose in and of itself.

[10] Including material from additional interviews and site visits.

utilizing Facebook360 spatial workstation plugin. The Facebook360 spatial workstation plugin within Pro Tools HD allows me to place a sound source in a 3D space that includes the X and Y axis (width and height). Some of the other audio sources were recorded both in mono and stereo with directional and binaural microphones. phones.

The final iteration can take up to several days of remixing, experimenting, and placing audio sources into spatial realms. This process of 3D sound mixing illuminates and enhances the whakaaro *(thinking)* and kōrero *(talking, speaking)* of each wāhine. It also refines timbre, dynamics and texture. When encountering a portrait in a 3D audio environment, the listener experiences an immersive sonic representation. We hear breathing next to our ear, traffic driving through our body and the intimacy of laugher. Here I am concerned with a holistic presence, the essence of being and sonic placement such that the portrait plays inside the listener.

6. THE AUDIO PORTRAITS OF WĀHINE MĀORI

The audio-portraits depict three wahine Māori for whom music is a central part of their being. Moana Maniapoto, Ramon Te Wake and Dr Te Rita Bernadette Papesch come from very different worlds that they navigate in distinctive ways. However, they are all connected by a central thread; these are Māori women for whom Mātauranga Māori is a way of knowing and understanding everything visible and invisible that exists in the universe. These portraits are an assertion of mana wahine and they exist as distinctly Māori expressions.

The portrait is composed entirely of sonic elements. Thus, sound becomes the artist's palette with the sound design becoming the compositional experience and canvas. A listener's experience is such that they hear, feel and sense the essence of the wāhine in an audio-portrait that is composed within a 360-sound environment of simultaneous sound cues. Each portrait is between five and seven minutes in duration, each should be played with headphones of best quality and preferably in a darkened space.

Provided here are the audio sketch, audio draft, hypothesis test and final audio portrait of Moana Maniapoto. In addition are the final audio portraits of Dr Te Rita Bernadette Papesch and Ramon te Wake.

Moana Maniapoto

Iwi (Tribal) affiliations: Ngāti Tuwharetoa, Tūhourangi, Ngāti Pikiao

Audio sketch:

https://soundcloud.com/maree-sheehan-1/moana-maniapoto-tahi-audio-sketch1/s-9YPqL?si=39209397b1264b538e44e9a3874590d4

Audio draft:

https://soundcloud.com/maree-sheehan-1/moana-maniapoto-audio-portrait-draft-1/s-zPgu7?si=6d11652d572a422d84707ea34bc4cee1

Hypothesis test 1:

https://soundcloud.com/maree-sheehan-1/moana-maniapoto-hypothesis-test-audio-portrait-draft-1/s-TwUvd?si=4578cbc7fe4f4bcc825e3032fa66a397

Hypothesis test 2:

https://soundcloud.com/maree-sheehan-1/moana-maniapoto-audio-hypothesis-test-2-comparative-final/s-0PJXu?si=2bacde5c19184964a20c178b19eb19c5

Final audio portrait:

https://soundcloud.com/maree-sheehan-1/moana-maniapoto-sonic-portrait/s-99LJU?si=19acb04c4b404352b857954be8411f05

Dr Te Rita Bernadette Papesch

Iwi (Tribal) affiliations: Ngāti Apapkura, Ngāti Maniapoto, Waikato, Ngāti Porou

Final audio portrait:

https://soundcloud.com/maree-sheehan-1/te-rita-papesch-audio-portraiture-final/s-eMzpS?si=e68f17317756406ab50ac01a924041ba

Ramon Te Wake

Iwi (Tribal) affiliations: Te Rarawa, Ngāti Whatua

Final audio portrait:

https://soundcloud.com/maree-sheehan-1/ramon-te-wake-final-audio-portrait/s-gY8bK?si=228823e437094bcebe7a121c2dba50ab

This paper provides an understanding of my creative processes of reflective practice, iterative development of audio portraiture portraiture through three forms of experiments, audio-sketching, audio-drafting and refined hypothesis testing.

The creative and analytic procedures (in the interpretation of the 'data' provided by the protagonists of the audio portraits) are further analyzed and understood in my full PhD thesis which can be found in the references. Furthermore, the specifics of the spatialization, immersive and binaural practices and its contribution to the audio portraits are discussed.

7. CONCLUSIONS

Employing a Kaupapa Māori paradigm has provided a useful research orientation that utilizes familiar cultural values. These values have proven effective because I am a Māori woman working with other Māori women, but at the same time, my research draws upon the knowledge of non-Māori contributors. Tikanga has provided protocols and processes that support Māori values in research. Of these values, manaakitanga[11] has been a guiding light, orienting the manner, in which I have developed relationships with each wāhine and her whānau *(family).* This value was extended into the respectful and appreciative way.

In the audio portraits, space is not only physical or chronological, it is also esoteric. Because these are portraits of three unique Māori women, there is no attempt to separate the spiritual and secular worlds [55, 56, 57, 58]. Accordingly, I sought to depict in the sonic design of the portraits, the wairua of the wahine. The expression of wairua through binaural sound cues required sensitive negotiations between sound sources, their spatial positioning and the listener's perception.

An audio-portrait can be experienced as having mauri that is distinct from a depiction of wairua. The mauri is the essential living vibration of the work; the essence of what reaches out to the life force of the listener and calls it into an intimate relationship. This life force is expressed through the blending of sound frequencies that move vibrationally. This vibration creates a movement, which can be likened to a tidal ebb and flow.

A reflective practice approach provided a framework that was integral for both iterative development and diverse forms of thinking and rethinking. The three forms of experimentation: sketching, drafting and refining the portraits (through hypothesis testing), extended the way I have worked as a musician in the past. They enabled me to be more interrogative because they elevated the importance of time, risk taking and critical reflection. The adoption of an artistic paradigm meant that sound technological knowledge operated in a highly creative arena. Given that many of the emerging sound technologies had not been utilized in art practice, the research opened up new pathways and pushed boundaries that suggest rich and exciting potentials for future applications.

Given that portraiture of wāhine has remained largely confined to concerns with pictorial imagery, I demonstrate how representations might artistically draw on the potentials of a rich spectrum of purely aural modes that are integral to Māori ways of expressing knowledge and identity.

Acknowledgments

It has been an honor to be in the company of Dr Te Rita Papesch, Moana Maniapoto and Ramon Te Wake. I wish to thank each of them for sharing their lives with me, along with their insights, generosity of spirit and willingness to be a part of this study.

Appreciation and thanks to my Primary Supervisor Professor Welby Ings and secondary supervisor Hinematau McNeill, Te Ara Poutama; the School of Māori and Indigenous Studies, and the School of Art and Design. Auckland University of Technology.

8. REFERENCES

[1] H. M. Mead. *Tikanga Māori: Living by Māori Values*. Huia Publishers, 2003.

[2] M. Marsden and T.C Royal. *The Woven Universe: Selected Writings of Rev. Māori Marsden*. Ótaki, New Zealand: Estate of Rev. Māori Marsden, 2003.

[3] P. Moon. *The Tohunga Journal: Hohepa Kereopa, Rua Kenana, and Maungapohatu*. David Ling Publisher, 2008.

[4] B. Boren. "History of 3D Sound." In A. Roginska and P. Geluso (Eds.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (pp. 40–62). Routledge. 2018.

[5] M. Sheehan, *The Sound of Identity Interpreting the Multi-Dimensionality of Wāhine Māori Through Audio Portraiture* (doctoral thesis, Auckland University of Technology). 2020. Tuwhera Open Access Online. https://hdl.dandle.net/10292/13262.

[6] L. Candy. *Practice Based Research: A Guide.* Sydney: University of Technology, Creativity & Cognition Studios, 2006. http://www.creativityandcognition.com. 2006.

[7] R. Bishop, *Collaborative Research Stories: Whakawhanaungātangā*. Dunmore Press, 1996.

[8] F. Cram. "Rangahau Māori: Tona Tika, Tono Pono – The Validity and Integrity of Māori Research." In M. Tolich (Ed.), *Research Ethics in Aotearoa New Zealand – Concepts, Practice, Critique* (pp. 35–52). Pearson Education, 2001.

[9] L. T. Smith. "Māori Women: Discourses, Projects and Mana Wahine." In S. Middleton and A. Jones (Eds.)*, Women and Education in Aotearoa,* (pp. 33–51). Bridget Williams Books Limited, 1992.

[10] L. T. Smith. *Decolonizing Methodologies: Research and Indigenous People.* London: Zed Books, 1999.

[11] D. A. Schön. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, 1983.

[12] K. Niedderer and S. Roworth-Stokes. *The Role and Use of Creative Practice in Research and Its Contribution to Knowledge.* IASDAR07, International Association of Societies and Design Research. University College for the Creative Arts, United Kingdom, 2007.

[13] G. W Davidson, M. A Seaton, and J. Simpson. *Chambers Concise 20th Century Dictionary*. W & R Chambers, 1985.

[14] Woodall. *Portraiture: Facing the Subject*. Manchester University Press, 1997.

[15] R. Brilliant. *Portraiture*. Reaktion, 1991.

[16] C. H. Claudy. "The Modern Idea in Photographic Portraiture." *Art and Progress* 6 (7), 1915, pp. 231–235. https://www.jstor.org/stable/20561452.

[11] Manaakitanga – "the show of respect, generosity, kindness" [29, p. 172].

[17] W. Leister. "Mona Lisa on a Bad Day: Or the Impossibility of Neutrality and the Non-Likeness of Photographic Portraiture." *Photography & Culture* 3(2). 2010.

[18] N. Yablon. "Posing for Posterity: Photographic Portraiture and the Invention of the Time Capsule, 1876–89." *History of Photography* 38(4), 2014, pp. 331–355.

[19] G. S. Baylis. "The Photographic Portrait: A Means to Surveillance and Subversion." *Early Popular Visual Culture*, 16(1), 2018, pp. 1–23.

[20] D. Schiffrin. "Narrative as Self-Portrait: Sociolinguistic Constructions of Identity." *Language in Society, 25*(2), 1996, pp. 167–203.

[21] S. Lawrence-Lightfoot and J. H. Davis. *The Art and Science of Portraiture*. Jossey-Bass, 1997.

[22] C. Orwig. *Narrative Portraiture: Foundations of Portraiture,* 2011. [Online video]. https://www.lynda.com/Photography-Photo-Assignments-tutorials/Narrative-Portraiture-Foundations-of-Portraiture/73663-2.html.

[23] L. Reihana and R. Devenport, *Digital Marae*. Govett Brewster Art Gallery. 2009.

[24] S. Cubitt. *The Practice of Light: A Genealogy of Visual Technologies From Prints to Pixels*. Cambridge: MIT Press, 2014.

[25] E. Altintzoglou, *Portraiture and Critical Reflections on Being*. Routledge, 2018.

[26] L. T. T. R. Smith, "Kaupapa Māori Research." In M. Battiste (Ed.), *Reclaiming Indigenous Voice and Vision.* University of British Columbia Press, 2000, pp. 225–247.

[27] A.Yates-Smith. *Hine! E Hine! Rediscovering the Feminine in Māori Spirituality* (Unpublished doctoral thesis). University of Waikato, Hamilton, New Zealand, 1998.

[28] P. M. Ryan. *The Reed Dictionary of Modern Māori*. Reed, 1995.

[29] H. W. Williams. *Dictionary of the Māori Language* (1st Ed). GP Publications, 2003.

[30] P. Miller. *Sound Unbound: Sampling Digital Music and Culture*. MIT Press. 2008.

[31] I. Friðriksdóttir. *Portrait: Expression Through Sound, Representation and Interpretation* (MA thesis, Mills College, Oakland, California). [Researcher copy]. 2017.

[32] B. S. B. Ndikung. *Where Do We Go From Here: For They Shall Be Heard*. Frieze, 2018.

[33] K. Gridley. *Sound Portraits*. 2019. Retrieved from http://www.kategridley.com/home-2/sound-portraits/.

[34] D. Sadja. *Sound Portraits Radio — Experimental Electronic Music Radio*. Retrieved 31 July 2019 from http://www.soundportraits.info/.

[35] L. Bongiovanni. *Music Portrait: The Sound of You*, 2018. [Online video]. https://www.yourmusicportrait.com/#Product.

[36] J. S. Walden. "Composing Character in Musical Portraits: Carl Philipp Emanuel Bach and L'Aly Rupalich." *Musical Quarterly*, *91*(3–4), 2008, p. 379.

[37] D. Toop. *Sonic Boom Exhibition Catalogue*. Hayward Gallery, 2000.

[38] A. Licht. *Sound Art: Beyond Music, Between Categories*. Rizzoli. 2007.

[39] A. Bhaugeerutty. *Sound Art: Discourses of Definition in the Contemporary Artworld.* Doctoral Thesis, Newcastle University, Newcastle, England, 2017.

[40] C. Cox and D. Warner. *Audio Culture; Readings in Modern Music*. NY: Continuum. 2004.

[41] V. Algazi and R. Duda. "Headphone-Based Spatial Sound." *IEEE Signal Processing Magazine*, 28(1), 2011, pp. 33–42.

[42] A. Roginska and P. Geluso. "Introduction." In A. Roginska and P. Geluso (Eds.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio.* Routledge, 2018.

[43] H. Møller. "Fundamentals of Binaural Technology." *Applied Acoustics*, *36*(3–4), 1992, pp. 171–218.

[44] S. Scrivener. *Reflection in and on Action and Practice in Creative-Production Doctoral Projects in Art & Design: The Foundations of Practice-Based Research.* Working papers in art & design. (Original work published by University of Hertfordshire, Centre for Research into Practice), 2000.

[45] D. Boud, R. Keogh, and D. Walker. *Reflection: Turning Experience into Learning.* NY: Nichols, 1985.

[46] P. McClure. *Reflection on Practice*. Routledge, 2005.

[47] S. Høyrup and B. Elkjaer. "Reflection: Taking it Beyond the Individual." *Productive Reflection at Work,* pp. 29–42, Research Gate, 2006.

[48] W. Ings. *Talking Pictures: A Creative Utilization of Structural and Aesthetic Profiles From Narrative Music Videos and Television Commercials in a Non-Spoken Film Text* (Doctoral thesis). Auckland University of Technology, 2006.

[49] B. G. Douglass and C. Moustakas. "Heuristic Inquiry: The Internal Search to Know." *Journal of Humanistic Psychology* 25(39), pp. 39–55, 1985.

[50] S. Sela-Smith. "Heuristic Research: A Review and Critique of Moustakas's Method." *Journal of Humanistic Psychology*, *42*(23), pp. 53–88. 2002.

[51] W. Ings. "Managing Heuristics as a Method of Inquiry in Autobiographical Graphic Design Theses." *The International Journal of Art &* Design *30* (2), pp. 226–241. 2011.

[52] R. R. Pere. "Te Wheke: Whaia te Maramatanga me te Aroha." In S. Middleton (Ed.), *Women and Education in Aotearoa* (pp. 6–19). Allen & Unwin, Port Nicholson Press. 1988.

[53] M. Csikszentmihalyi. *Flow: The psychology of optimal experience.* Harper & Row, 1990.

[54] M. Csikszentmihalyi. *The systems model of creativity: the collected works of Mihaly Csikszentmihalyi*. Springer, 2014.

[55] C. Barlow. *Tikanga Whakaaro: Key Concepts in Māori Culture*. Oxford University Press. 1991.

[56] D. Brown. *Māori Architecture: Whare Māori*. 2014.

[57] A. Salmond and A. Hui. *A Study of Māori Ceremonial Gatherings*. A.H. & A.W. Reed, 1975.

[58] R. Walker, *Ka Whawhai Tonu Matou: Struggle Without End*. Penguin, 1990.

A MATTER OF NOTATION: CASE STUDIES IN MATERIALITY, TEMPORALITY, AND THE SCORE-OBJECT

Kate Milligan
Independent Researcher
kate@milligan.net.au

Mattering is simultaneously a matter of substance and significance, most evidently perhaps when it is the nature of matter that is in question, when the smallest parts of matter are found to be capable of exploding deeply entrenched ideas [1, p. 3].

– Karen Barad

The art or cultural object is endowed with a capacity to generate both time and space [2, p. 368].

– Georgina Born

ABSTRACT

From a posthuman perspective, music-making can be understood as a lively assemblage of human and non-human agents. Within this framework the term 'music culture' expands to include more-than-human contexts, scales, and temporalities. The question therefore arises: how might composers facilitate musical experiences "beyond the 'mesoscopic bubble', within which our human perception and phenomenological considerations are still enclosed" [3, p. 1]? In this paper, I position the score-object as a subject with material vitality, and an active participant in the balance of agencies in the musical assemblage. Specifically, I employ the notion of "medium time" as outlined by Georgina Born to understand how musical objects act independently in and on time.

Through material intervention in my notational practice, I assess the balance of these agencies as they arise in musical experiences. More specifically, in this material context, I am continuing an investigation of "watery logics", incorporating water into my practice as medium (and subject) of primary focus. In this paper, I present two material-compositional case studies from 2023, wherein animated notation is re-mediated by water for generative visual results, resulting in fluid dispersion of musical agency. Finally, I position this discussion within a broader posthuman discourse of ethics.

1. INTRODUCTION – TIME AND OBJECTHOOD

Musicians trained within the Western Classical idiom know intimately the compulsion for score study, and the associated faith in icons, documents, and visual motifs inherited from our forebears. We love to pore over scores – a living High Romantic practice. In 2016, Mahler's handwritten Second Symphony sold for £4.5 million, and remains the most expensive score ever sold at auction [4]. How we revere these material fragments of history.

Traditional customs in Western Art Music often cite the manuscript as a source of essential compositional knowledge – as Taruskin wryly put it, the "inviolable musical 'object'" [5, p. 339]. The sanctity of the urtext and the politics of reproduction – or, the material authenticity and transformation of the score through time – still govern performance practices around the world. Beginning in the late 20th century, a lineage of music philosophers has challenged these material-temporal ideologies, culminating in contemporary, post-Deleuzian discourse that understands the past-future relationship (and the role of a musical object, or score, within this) as one of deep complexity: "The object is changing, the environment where it is posited is changing, and the subject-receiver is changing" [5, p. 72].

De Assis offers a *logic of assemblage* for this revitalised epistemology, which serves to describe the dynamic systems arising from entanglements of human and more-than-human objects and practices [5, p. 99-100]. The effect of this logic is "the overcoming of unity in favour of multiplicity, of essence in favour of event, of being in favour of becoming". The question therefore arises: how might we understand the new role of the score within the musical assemblage, and what becomes of the composer against the backdrop of inherited ideologies in Western Art Music?

In contemporary practice, increasing numbers of composers are concerned with posthuman issues, and the question of human subjectivity in an ecological framework – a flat ontology aligned with that of the assemblage. Furthermore, many composers are engaging with temporality in this context, for example Liza Lim *(Multispecies Knots of Ethical Time*, 2023) [6], John Luther Adams *(An Atlas of Deep Time*, 2021) [7], and Harrison Birtwistle *(Deep Time*, 2016) [8].

Of these composers (a non-exhaustive list) and within this framework of eco-logic, Lim makes an explicit connection to the *agency of the score*: "What levels of sensory abundance could be unlocked for the performer if they engage with the sentiency of notation?.. One can argue things into lifelessness, or one can enter into a weaving, rhythmic, multi-modal sensory relation where what one does is continuous with a wider living world" [9]. She argues for experiences of posthuman time that come about by the very materiality of music-making. With reference to her work *The Tailor of Time* (2023), she asks: "How does one tie 'knots' in time, sew 'pleats' or 'pockets' in time and not only disrupt but rip up, create rifts in, and erode the materiality of time?" [9].

It is here that I situate my work. My compositional practice takes a material approach to the associated practices of notation and scoring, employing graphic, animated, and physical media. I am interested in the score as a trans-temporal object – that which transmits cultural information through time, presents dynamically in different temporal contexts, but retains some sense of objecthood agency so as to be partially independent of the human cultural context. My own notation playfully engages with theories of temporality, often with reference to anthropologies of future-seeing and prophecy.

2. NEW MATERIALISM AND THE POSTHUMAN

Before delving too far into the new materialist epistemology that furnishes this understanding of the agentic score-object, I offer first a brief discussion of the posthuman – the ethical foundation and motivating force for my compositional practice – and some context on the idea on the musical assemblage. In this paper, I take an explicitly posthuman stance to view music-making as a lively assemblage of human and non-human, living and non-living agents, one of which is the score. Karen Barad offers the following:

> Posthumanism... is about taking issue with human exceptionalism while being accountable for the role we play in the differential constitution and differential positioning of the human among other creatures (both living and non-living). Posthumanism does not attribute the source of all change to culture, denying nature any sense of agency or historicity. In fact, it refuses the idea of a natural (or, for that matter, a purely cultural) division between nature and culture, calling for an accounting of how this boundary is actively configured and reconfigured [10, p. 136].

In this definition I read accountability, empathy, and reflexivity – the ongoing, species-wide evaluation of self and subjecthood, the probing of our entanglements with the other, and a practice which I believe is enacted inherently (knowingly or unknowingly) within the musical assemblage. A term with a rich discursive history [11], 'assemblage' as I understand it serves to describe the dynamic and emergent processes in ways of being together: "Assemblages are open-ended gatherings. They allow us to ask about communal effects without assuming them. They show us potential histories in the making... How do gatherings sometimes become 'happenings', that is, greater than the sum of their parts?" [12, pp. 22–23]. A musical assemblage is a profound network of relationships between living (composers, performers, conductors, audiences, human microbiomes) and non-living agents (instruments, venues, acoustics, weather systems, and, of course, scores). Assemblage theory articulates the dance of agencies that occurs across the various temporalities enlivened by these agents.

In the quotation above Anna Tsing alludes to the socio-political power of the assemblage, the potential energy in the act of becoming together across a curious scale of time flipped back upon itself: "histories in the making". Barad also positions posthumanism within a larger cultural-ethical framework, with very specific discursive treatment of the present "the becoming of the world is a deeply ethical matter" [10, p. 185]. If ethics were to exist in a medium, or to become material, surely it would be in the weave of time. Our values and morality exist inherently as negotiation of past and future – that which we judge to be good must have some precedent, and, of course, consequence – and the liveliness of the musical assemblage compels us to reckon with the ethics of agency in real-time. This negotiation is felt in the subtleties of ensemble musicianship, and in the vibrancy of music-making as a temporal-material medium.

Posthuman and assemblage theory promise greener pastures for many areas of musicology. In this paper I am concerned specifically with the vitality of non-living matter in the musical assemblage, and the role of scores as trans-temporal objects within this framework. Returning to Mahler's Second Symphony manuscript, and the strange power it held over the buyer who decided to drop a substantial amount of cash to secure ownership. For those similarly enchanted, Jane Bennet compels them to "linger in those moments during which they find themselves fascinated by objects, taking them as clues to the material vitality that they share" [13, p. 17]. There is something about the manuscript's objecthood – it's material fragility, coupled with its remarkable journey through time, a trajectory quite exempt from any singular human frame of reference – that moves us. As the score was once made, it continues to make, playing a vital part in the evolution of living music cultures, not least historical music practices.

The epistemology offered by new materialism constitutes a radical renegotiation of the role of the score. Within this framework of music-making, "non-human materialities [are] bona fide participants rather than... recalcitrant objects, social constructs, or instrumentalities" [14, p. 62]. Where post/structuralist critics may challenge this, bearing arms with the notion of representation – scores are, after all, composed of iconography, and we speak so often of the separate process of *interpretation* as the humanistic organisation of these codified symbols – Barad offers the following material reconciliation of signifier and the signified: "The common-sense view of representationalism – the belief that representations serve a mediating function between knower and known – ... displays a deep mistrust of matter, holding it off at a distance, figuring it as passive, immutable, and mute, in need of the mark of an external force like culture or history to complete it" [10, p. 133]. That is to say: "Matter and meaning are not separate elements" [1, p. 3]. The material choices we make as composers in our notation matter, and cannot be divorced from the meaning gleaned by those who read it. A material approach to the score necessitates deeper deliberation on mediation itself, and calls to question what musicality and music-making might evolve to mean in different material contexts.

3. MATERIAL PRECEDENTS IN THE ARTWORLD

I am interested specifically in musical-temporal phenomena produced by experiments in notational materiality. Here, I take methodological cues from the artworld, wherein discussions of materiality and meaning have long been foregrounded, and whereby the reverence of material artefacts has no direct equivalence. Choosing exactly which material to use in my experiments has been a process grounded in research, and largely informed by an ancestry of artists and makers working with themes of music, time, and objecthood. The case studies of my own material approaches, described below, build on work by the likes of Christian Marclay, Cornelia Parker, and Man Ray – artists who have engaged conceptually with theories of temporality and music-making.

May Ray's *Object to be Destroyed* (1923) [15], a readymade, takes what is perhaps the most widely recognised symbol of regimented musical time – the metronome – and instils it with a violent potential energy. In the work, a cropped photo of a woman's eye (belonging to Lee Miller, his ex-lover) is affixed to the metronome's swinging hand – a human-object cyborg companion in his studio, that which would observe his artistic process, ominously dictating a timeline of its own. Here, Man Ray challenges the duality of subject and object with explicit reference to the audience gaze. Where previously a human audience plays a passive role in observation, Man Ray's *Object to be Destroyed* asserts an active role in the creation of artwork in a dynamic embodiment of more-than-human time.

Upon his split with Miller in 1932, Man Ray recreated the work for the first time with the following text instruction (a score of sorts): "Cut out the eye from a photograph of one who has been loved but is seen no more. Attach the eye to the pendulum of a metronome and regulate the weight to suit the tempo desired. Keep doing to the limit of endurance. With a hammer well-aimed, try to destroy the whole at a single blow" [15]. This was the first of a series of manifestations that would occur as the object was destroyed and re-made many times over. Like a work of music, the artist's initial material model and the accompanying instruction served as a score for subsequent performative iterations over the 20th century. The work assumed various titles, including *Indestructible Objects*, and *Perpetual Motif*, alluding to the cycle of destruction and regeneration that would out-live the artist himself, and his original audiences.

Where Man Ray's object is subject to a violent, perpetual cycle of life and death, Cornelia Parker's *Perpetual Canon* (2004) [16], is destruction frozen in time. Composed of defunct brass instruments – tubas, trombones, trumpets, a sousaphone – squashed flat by an industrial press, the work is suspended in a circle organised around a lightbulb, which throws a shadow of the instruments on the surrounding walls. Parker is interested in the material memory of her work, and uses objects with significant cultural-historical trajectories, in this instance the British Brass Band. The title refers at once to the canon as musical form – the work a material representation of infinite repetition – and perhaps also the notion of circular breathing, evoking a visceral response in her audience of immeasurable breathlessness. *Perpetual Canon* presents boundless time in a suspended moment of quiet destruction, forever recreating the moment where her materials were crushed flat.

There is a something about the score-object that presents to us a perpetual motif, or a perpetual canon of reoccurrence in its very materiality. This is the quality of transtemporality that I refer to in my own work, a curious distortion of linear time: "To say that art is taken out of time is not to say it is timeless but that its relationship to time and history becomes multiple" [17, p. 23]. In the final example I will draw from the artworld, Christian Marclay distorts musical time with his visual-material practice in *Drink Away* (2014) [18]. He superimposes crown glass (the product of a specific British glass-blowing technique, often to be spotted in pub windows) on sheet music of an English drinking song. The diffraction of light through the glass medium bends the linearity of the musical stave, speaking not only to the sense of expanded musical time, but also to experiences of drunken, distorted time. It is from *Drink Away* that I take material cues for my own notation experiments (explained below), beginning with representation of the musical stave as the most prominent Western icon for musical time.

4. MEDIUM TIME

Anthropological musicologist Georgina Born notes that "music provides an auspicious terrain for retheorising time and history" [2, p. 362]. According to Born, music offers fertile ground for thinking through past, present, and future, not only for its phenomenological and experiential characteristics as a lively temporal medium, but also for the host of temporal ideologies inherited by practitioners of the Art Music genre: "music produces time through the contingent articulation of its several temporalities, while in turn the variant temporalities immanent in social, cultural, political, and technological change mediate the evolution of music and musical genres" [2, p. 371]. To think of the *production* of time – the tantalising idea that somehow time itself might be manipulated in arts practice – relocates this discussion from the realm of physics and science fiction, and calls to question the ethics of *making* time in musical assemblages. What kind of time is made, and by whom?

Scaffolded by the interdisciplinary discourse of temporal philosophy, Born identifies four musical temporalities, which provide a neat framework for theorising time in musical contexts:

1. The first temporality is well recognised: the qualities of temporal unfolding of musical sound as it enlivens musico-social experience and "entrains" musical attention...
2. A second temporality [is derived from] ... the radically object-centered, posthumanist perspective that any musical object or event itself animates a temporality through its retentions and protentions – through connections to prior and prospective objects or events...
3. A third temporality concerns the variable temporalities produced by particular genres as themselves objects distributed in time, in the guise of any genre's characteristic metarhythms of repetition and difference...
4. The fourth returns to the human via temporal ontologies: encultured ways of living and conceiving time [2, pp. 372–374].

Where the first and fourth temporalities are perhaps the most widely appreciated in musicology, the third is typical of Born's longstanding engagement with genre theory. It is, however, the second of Born's temporalities that interests me for two reasons: firstly, the theory's posthuman motivation, and the world of temporal philosophy that is revealed upon stepping out of the human mesoscopic bubble, and secondly; the creative possibilities for temporal intervention (the above-mentioned *making* time) as a composer presented with new material and object-oriented pathways. As Tsing writes regarding more-than-human time-making: "The curiosity I advocate follows such multiple temporalities, revitalising description and imagination" [12, p. 21].

Born elaborates on "the temporality of the medium" as "a temporality that in each case interferes technically, conceptually, and aesthetically with the musical temporalities at issue" [2, p. 380]. If we are to think of mediation as interference, then we ascribe agency or intent to the medium actor. And once the discussion returns to agency, we must necessarily return to the ethical question that is implicit in the balance of agencies within any musical assemblage. What kind of time is made, and by whom? Writing from the human-scale reference point, it is easy to identify hegemonic temporalities – the temporality of progress, of neoliberal time, of Museum Culture, the religious temporalities of fate and determinism, to name a few – at play within Western culture spheres. By the simple virtue of offering more-than-human temporal alternatives, the creative possibilities for sustainable coexistence (for example, along deep, post-anthropocentric timelines) blossom. Whilst this might sound like an extrapolation upon theories of medium time, I argue that just the seed of the posthuman idea – and material intervention – has immense implications for musical practice. As Barad puts it: "the smallest parts of matter are found to be capable of exploding deeply entrenched ideas" [1, p. 3]

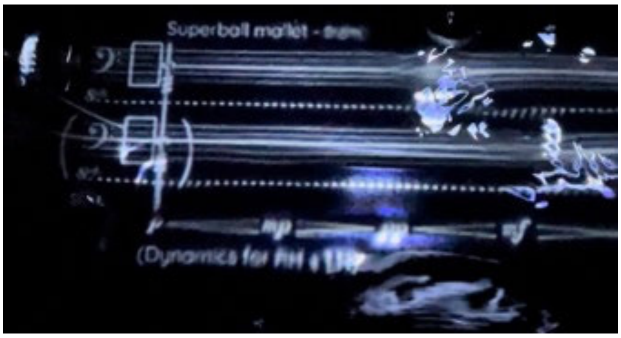
5. EARLY SKETCHES / WATERY TIME

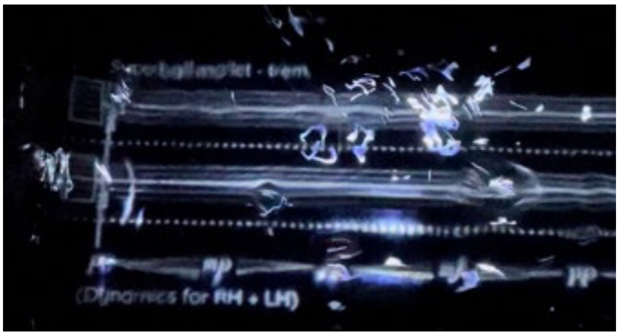
My compositional methodology aims to play with the temporality of the medium, testing different types of materiality against goals of reformed, posthuman temporal understandings in my audiences. In this paper, I present two case studies of my notational approach within the genre of *performance-installation*. As this research is ongoing, at the time of writing I have no measurable results related to reformation in audience perspective. Rather, findings are derived from object-oriented observations and the dynamics of the performing ensemble. The works outlined in this paper serve as a starting point, the beginning of a new pathway of practice-based inquiry that I hope will evolve into a much larger research project. The outcome will be to observe "the multiple temporalities produced by the musical object" and measure the degree to which "they are informed by, and as they may drift from, the temporal ontologies of human actors, and as they interact with other heterogeneous trajectories of historical change" [2, p. 380].

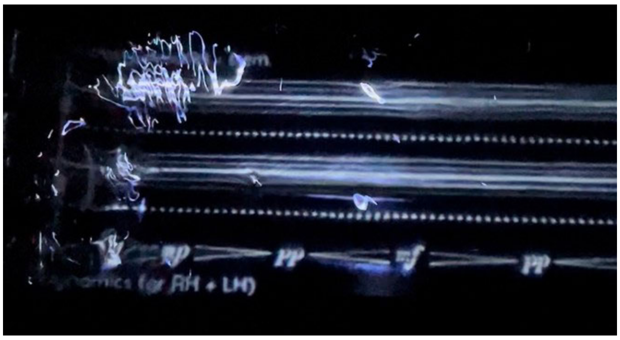
Experimentation (see Figures 1 through 8) began at a small-scale, prototype level in late 2022, with speculation as to how posthuman understandings of notation might be communicated to lay audiences. Perhaps easier than proving the inherent agency of a static object is observing agency in livelier nonhuman systems or materials. That is to say, where Barad and Bennet would transparently see a networked exchange of more-than-human forces even in the most canonical of music-making, the lay audience may struggle without the appropriate philosophical furnishing. In my compositional practice, I therefore refer to particularly dynamic non-human materials to more clearly demonstrate to an audience shared compositional agency. Perhaps for its ubiquity, I chose to use *water* as a material intervention in notational practice. What kinds of more-than-human creativity lie dormant in watery systems? How might we think about the past, present, and future across watery timescales?

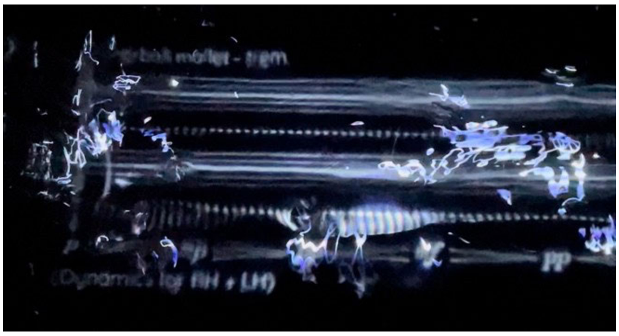
These sketches together demonstrate a variety of what I have come to call *water graphics* – ephemeral, emergent graphical output from a digitally animated input and watery mediation. Each of these approaches constituted a different treatment of temporal linearity as represented by the stave, and the graphical results show distinct material characteristics of the water. I was particularly drawn to the results of Experiment #3, whereby the stave became curved according to the circularity of the vessel, and the dynamism of the distortion was entirely determined by the relative position of the viewer. As a purely compositional or graphical exercise, these experiments put forward a series of interesting provocations for the interpretation of the hypothetical performer.

"We mix language, gods, bodies, and thought with water to produce the worlds and selves we inhabit" [19, p. 3]. Here Jamie Linton refers not only to global anthropological understandings of water, but also the productive power

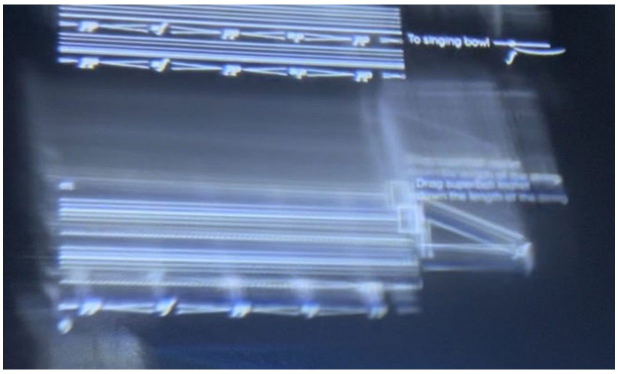


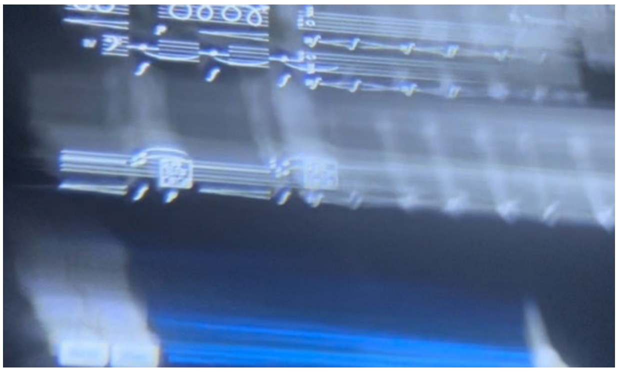




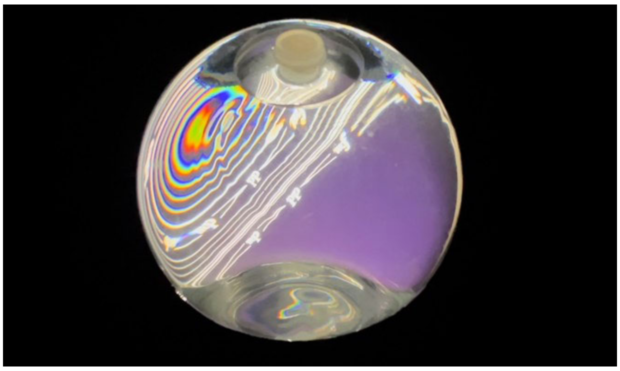


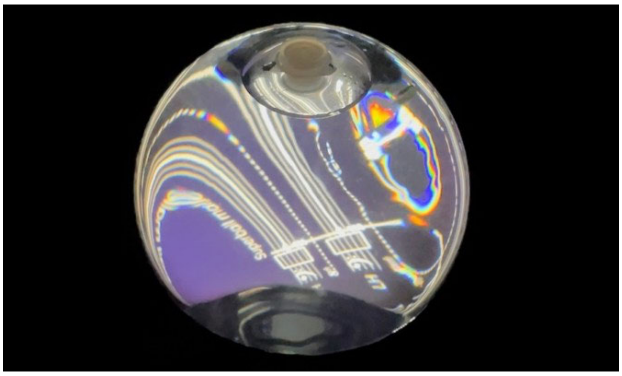
Figures 1, 2, 3, and 4. Experiment #1 from late 2022. Scrolling notation (using the Decibel ScorePlayer[1]) projected *onto* water and a matte black vessel. These images are stills from video footage—capturing ephemeral light phenomena as they are reflected off the water's restless surface.





Figures 5 and 6. Experiment #2 from late 2022. Scrolling notation projected *through* water and a glass vessel. These images are stills from video footage—capturing the distortion of the notation as the light source is shifted.





Figures 7 and 8. Experiment #3 from late 2022. Scrolling notation projected *through* water and a spherical glass vessel. These images are stills from video footage—capturing the distortion of the notation from different angles as the viewer moves around the sphere.

1 https://decibelnewmusic.com/decibel-scoreplayer

of water to impact society. He proposes an understanding of water as a dynamic process, that which might not be easily defined or understood in essential terms, but nevertheless with fluid material origin: "When we do contrive to slow down the flow for long enough to substantiate it in language, represent it in numbers, or confine it in Euclidian spaces, water transform and slips into impermanence" [19, p. 4]. A similar understanding of fluid logic is presented by the hydro-feminist scholar Astrida Neimanis, who recognises the conceptual usefulness of fluidity – particularly in a poststructural epistemological context – but nevertheless insists on the materiality of water as fundamental, lest discourse detach from abject, material realities: "I am not interested in fluidity and watery logic *only in the abstract*... It seems important that we pay attention to the specific ways in which water travels, and the specific kinds of bodies that certain waters comprise, transform, and dissolve" [20, p. 22].

So what can we learn from "think[ing] *with* water, and even learn[ing] *from* water" [20, p. 22] within the musical assemblage? Who or what stands to gain from centring lively watery processes in music-making? These are vast questions, with no immediate solution; only a longitudinal, collaborative, and multi-valent creative research process might go some way to answering them. In the meantime, I look to my three notational experiments for material clues.

Experiment #1 produced bright, ephemeral, and fragile graphics from the light reflected from the subtle movement of the water's surface. A process of generative scoring, these graphics bring new meaning to that which we understand as *animated:* they are truly lively, and intimately dependent on the balance of more-than-human agencies in any given performance space – the temperature, the air pressure, the collective breath. In their fugitivity, these graphics reconstitute the present moment. Experiment #2 made indeterminate the visual constitution of the notation with hazy, unfocussed graphics. Furthermore, this mediation made the notation multiple – the stave expanded vertically as individual lines were compounded, with obvious implications for range, repetition, and rhythm. The results of this experiment speak to the *multiplicity* of watery time. Experiment #3, as expressed above, curved the stave around the spherical vessel, and merged the notation with reflections like a visual palindrome. This experiment I feel has the greatest potential for experiments in musical form as determined by water, and subsequent investigation into alternative musical temporalities.

6. CASE STUDY #1 – *VISIONS | VESTIGES* (2023)

Lay understandings of the posthuman in my work are further fostered by deliberate visual-thematic choices. *Visions | Vestiges* (2023)[2] is a performance-installation for bass clarinet, cello, percussion, live electronics, and sculptural notation (glass, water, antique wood). In this work, the visual metaphor of a crystal ball speaks directly to commonplace understandings of object agency, temporality, and the more-than-human that many audiences (at least those across the Anglo-Celtic and European diaspora) will have been raised on since childhood. This work builds on the techniques of Experiment #3, whereby animated notation is distorted through the materiality of water and glass. The use of these materials, and their existing embeddedness in cultural context, contributes to the collaborative narrative of the performance work in which the audience is participant. Rather than dilute the posthuman message by scaffolding the work in metaphysical connotation, my hope is that reformed understandings of more-than-human agency in the musical assemblage will be made available to diverse, non-academic audiences.

Figure 9. *Visions | Vestiges* (2023). The notational sculpture was assembled from a Victorian-Era music stand, a handblown glass sphere, water, and an iPad displaying animated notation generated in the Decibel ScorePlayer.

This work constitutes the first of what I now call my *notational sculptures*, whereby notation is imbued with a heightened performative role within the musical assemblage. Assembled from a Victorian-Era music stand (that which was once used to store and display sheet music – the intended effect of which is a quasi-hauntological re-mediation of the past) and water encased within a handblown glass sphere, the crystal ball sculpture gives cultural, dramatic, and thematic context to the above-described graphic techniques of distorted musical linearity.

The passage of time, and the relationship between past and future are explored equally in the traditional compositional input – my own contribution a post-minimal approach to thematic material, structured with repetitive isorhythms that set each instrumentalist within their own individual temporality. The notation is designed to be read

[2] https://www.youtube.com/watch?v=HqMcqMWPrQg

Figure 10. *Visions | Vestiges* (2023). A performance work for bass clarinet, cello, percussion, and electronics. The notation is distorted through the media of glass and water, disrupting the temporal linearity of the stave.

directly from the crystal ball, and the curious temporal distortions are explored in real-time by the performers. Unfortunately, due to logistical restrictions at the premiere event, the three musicians were unable to simultaneously read the watery mediation for the premiere performance – a design failure that relegated the notational sculpture to a more passive role in the musical assemblage than originally intended.

7. CASE STUDY #2 – *THE FORECAST* (2023)

Crucial to discerning whether the watery materiality of my notational approach is, in fact, generating alternative musical temporalities is the performers' unimpeded access to the materials at hand. Reflecting on the logistical failures of *Visions* | *Vestiges* (the sculptural complement to the music-making was a key experiential factor for the audience, but the performers were estranged from this process despite being a lynchpin contingent of the musical assemblage) I set out to make another work that would have the musicians reading directly from watery graphics. *The Forecast* (2023) is a performance-installation for alto flute, viola, percussion (water gong, coin chimes, prepared vibraphone), spatial electronics, coins, and wishing well. An excerpt from my programme note reads as follows:

> *The Forecast* is a performance-installation wherein water is a co-creator. Fragments of text circulate in a custom wishing well, and instrumental whispers mingle with the shards of light reflected by the water's surface.
>
> The text is sourced from weather reports in old editions of *The West Australian* [newspaper]... Combined with the visual metaphor of a wishing well, these archival texts encourage us to think about humanity's historical entanglement with natural systems, and the methods we use – magical or scientific – to reconcile with these greater forces.

In this work, animated rings of text – the weather reports – are projected onto a shallow basin with a matte black

Figure 11. *The Forecast* (2023). Members of Decibel New Music Ensemble read generative water graphics from the wishing well. Photo by Edify Media.

surface, filled with water. Performers are not only required to interpret the text at an 'instrumental whisper' – a phonetic transcription to their idiom of their instrument – but also to interpret the ephemeral water graphics that are generated in real-time by oscillations in the surface tension of the water and the subsequent reflected light phenomena. This technique is reminiscent of Experiment #1, above. In *The Forecast*, I have provided the instrumentalists with a framework for interpreting these fleeting water graphics:

> When the surface of the water is still, the projected notation remains largely undisturbed. However, as the surface tension of the wishing well changes, flashes of light reflect off the water, creating ephemeral graphic notation.
>
> When these graphics occur *where you are reading your text* (on top of the text, or immediately adjacent to where you are reading), you respond. In these instances, the water graphics are read *instead* of the text, with no hesitation transitioning between the two. A specific technique (or in the percussionist's case, a specific instrument) is reserved – and must only be used – for the generative water graphics when they occur:
>
> i. Alto Flute: Whistle tones.
> ii. Viola: Natural and artificial harmonics.
> iii. Percussion: Coin chimes.
>
> Be guided by the aesthetic of the generative graphics – light, ephemeral, dynamic – in your improvisation.

As with *Visions* | *Vestiges*, this work furnishes the sociocultural context of the music-making with a folkloric reference that draws on themes of temporality and non-human agency – the wishing well. Again, the notation assumes a lively role in this performance work, plainly agentic in the eyes of the audience. The *visual liveness* of this generative, watery system meant that the performers' structured improvisation was strung with a present awareness shared by the audience. One member of Decibel described his experience of the temporality of this notation as "heightened", his relationship to the present moment reconstituted. Further experimentation is warranted to explore to full musical-temporal potential of these water graphics, and their impact on the balance of agencies within the musical assemblage. The implications of watery aesthetics in notation are yet to be fully determined. For now, *The Forecast* stands as my most developed integration of self-governing natural systems into notation, and my best attempt at a radical renegotiation of compositional agency in a posthuman frame of reference.

8. CONCLUSIONS

The complex connections between objecthood, time, and music-making are evident in the performance-installation outcomes of my research, whereby the score functions as a lively musical object in the present, but also exists as an artefact that moves in, on, and across past and future. Emphasis on the objecthood of the score facilitates a parallel emphasis on the materiality of the score, and how musical information is always mediated. Objecthood asserts the agency of the score in space, and makes tangible the trajectory of the musical object through time – a past and future that is separate from any individual human influence. In the case studies I have discussed in this paper, the score is a non-human actor in the musical assemblage.

The potential for headway to be made in notational technologies is significant when more-than-human logics and systems are given due acknowledgement. Within a new materialist framework, the role of the composer (as it stands amongst the scaffolding of inherited ideologies of Western Art Music) is disrupted. Unimpeded material access to the score-object – in this instance, water – for the performers and audience has proven to be a significant intervention in the musical assemblage. This research has scope to be greatly expanded in the future, particularly regarding methods of data gathering from performers and audiences, which has thus far been missing from my research design.

As I have mentioned periodically throughout this paper, my research is grounded in an ethical imperative – the complex negotiation of human and more-than-human agencies in music-making, and (more broadly) the politics of co-existence with vital materialities. I feel that is it important to speak with specificity, not only for the sake of an efficient research process, but also for the explicit communication of ethical discourse. Where water is concerned as the specific material subject of my investigation, a range of ethical imperatives arise: "Our 'making' of water as an imaginary is necessarily forged in the entanglement of our values with the very material matter at hand. It follows that our 'making' of water also includes all of the problematic ways we currently *re*make and *un*make it – as dirty, depleted, de-territorialised, for example" [20, p. 21]. Our treatment of water – the way we approach (or dismiss) the substance as an agentic and creative force – informs the limits of our creative potentials as a posthuman musical collective. In a precarious social, political, and ecological context, the subject of futurity is of existential concern. Material approaches to music notation, and the posthuman temporal learnings we might glean from them, might be a small yet important part of the solution.

Acknowledgments

Visions | Vestiges was commissioned for the London Symphony Orchestra's 2023 Soundhub Project.

The Forecast was commissioned by Decibel New Music Ensemble with assistance from The Digital Score Project funded by the European Research Council European Union's Horizon 2020 Research and Innovation programme.

Supported using public funding by
ARTS COUNCIL ENGLAND

9. REFERENCES

[1] K. Barad, "The Science and Ethics of Mattering," in *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham and London, UK: Duke University Press, 2007, pp. 3–38.

[2] G. Born, "Making Time: Temporality, History, and the Cultural Object," *New Literary History*, Vol. 46, No. 3, pp. 361–386, Summer 2015. https://doi.org/10.1353/nlh.2015.0025

[3] J. Hauser and L. Strecker, "On Microperformativity," *Performance Research*, Vol. 25, No. 3, 2020, pp. 1–7. https://doi.org/10.1080/13528165.2020.1807739

[4] Author Unknown. "Gustav Mahler £4.5m manuscript breaks record at Sotheby's." BBC News. https://www.bbc.co.uk/news/entertainment-arts-38142482 (accessed Nov. 28, 2023).

[5] P. de Assis. *Logic of Experimentation: Rethinking Music Performance through Artistic Research*. Leuven, BE: Leuven University Press, 2018.

[6] L. Lim. "Multispecies Knots of Ethical Time – who are we and who else is here?" Liza Lim: Composer. https://lizalimcomposer.com/2023/12/11/multispecies-knots-of-ethical-time-who-are-we-and-who-else-is-here (accessed Dec. 10, 2023).

[7] J. L. Adams. "An Atlas of Deep Time." John Luther Adams: Works/Orchestra. https://www.johnlutheradams.net/works/orchestra/7 (accessed Dec. 10, 2023).

[8] H. Birtwistle. *Deep Time*. Boosey & Hawkes. https://www.boosey.com/cr/music/Harrison-

Birtwistle-Deep-Time/100197 (accessed Dec. 10, 2023).

[9] L. Lim. "The Tailor of Time (2023): time paradoxes and the sentiency of notation." Liza Lim: Composer. https://lizalimcomposer.com/2023/10/13/the-tailor-of-time-2023-time-paradoxes-and-the-sentiency-of-notation (accessed Dec. 10, 2023).

[10] K. Barad, "Agential Realism: How Material-Discursive Practices Matter," in *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham and London, UK: Duke University Press, 2007, pp. 132–188.

[11] B. Latour, *Reassembling the Social, An Introduction to Actor-Network-Theory*. New York, USA: Oxford University Press, 2005.

[12] A. L. Tsing, "Arts of Noticing," in *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins*. Oxford, UK: Princeton University Press, 2015, pp. 17–26.

[13] J. Bennett, "The Force of Things," in *Vibrant Matter: A Political Ecology of Things*. Durham and London, UK: Duke University Press, 2010, pp. 1–19.

[14] J. Bennett, "Neither Vitalism nor Mechanism," in *Vibrant Matter: A Political Ecology of Things*. Durham and London, UK: Duke University Press, 2010, pp. 62–81.

[15] Author Unknown. "May Ray: *Indestructible Object*: 1923, remade 1933, editioned replica 1965." Tate. https://www.tate.org.uk/art/artworks/man-ray-indestructible-object-t07614 (accessed Nov. 28, 2023).

[16] Author Unknown. "Cornelia Parker: Exhibition Guide." Tate. https://www.tate.org.uk/whats-on/tate-britain/cornelia-parker/exhibition-guide (accessed Nov. 28, 2023).

[17] A. Nagel, "Learning to Live Without Artistic Periods,' in *Medieval Modern: Art Out of Time*. London, UK: Thames and Hudson, 2012, pp. 22–26.

[18] Author Unknown. "Christian Marclay, *Drink Away*, 2014." Artsy.net. https://www.artsy.net/artwork/christian-marclay-drink-away (accessed Nov. 28, 2023).

[19] J. Linton, "Fixing the Flow: The Things We Make of Water," in *What Is Water? A History of Modern Abstraction*. Vancouver: UBC Press, 2010, pp. 3–23.

[20] A. Neimanis, "Introduction: Figuring Bodies of Water," in *Bodies of Water: Posthuman Feminist Phenomenology*. London, UK: Bloomsbury Academic, 2017, pp. 1–26.

PERFECT INFORMATION: SCORES AS SELF-EVIDENT PROCESSES

James Saunders
Open Scores Lab, Bath Spa University, UK
j.saunders@bathspa.ac.uk

ABSTRACT

Self-evident processes in music suggest the possibility of close engagement with the operation of a piece, with the potential for empathic and communal experiences to be had by both participants and observers. Taking the notion of perfect and imperfect information in games as a starting point, the paper considers what might be gained by presenting scores as systems of perfect information. In games with perfect or public information, all the necessary information is made available to us, so we have the potential to understand player choices and empathise with them when spectating. In process music, if the necessary information is made available to us we also have the potential to understand player choices and empathise with them. The paper considers the development of self-evident music in this context, proposing three modes of information delivery: demonstrating, explaining and showing. These modes are considered in relation to the management of information flow and the corresponding cognitive load placed on listeners through the control of speed, density, simultaneity and sequentiality in the presentation of instructions.

1. INTRODUCTION

In a game such as chess, players have access to *perfect information*, where 'all players have complete knowledge about every element in the game at all times.' [1, p. 204] What is happening in terms of the movement and relationships between pieces is always visible, both to the players and, crucially, an audience. In a card game, such as whist, however, 'some of the game information may be hidden from players during the game' [1, p. 204], where managing such *imperfect information* might form part of the challenge. The parallels between games and other instruction-driven activities, such as scored music or instructional art, suggest the possibility of considering systems of perfect and imperfect information in those contexts to see how they impact on our experience of such work as participants or observers. Specifically, what might be gained by presenting scores as systems of perfect information?

While there are many examples of pieces which deal in perfect information in different ways, perhaps the most common form of presentation is that of imperfect information. For players in an ensemble context, the starting point may be that of not initially knowing what others are doing as they have access only to their own part, or there may be events that are more unpredictable to which they must react. For audiences, much of this working is also hidden, leaving them only with the trace of the process apparent in the performed result. There are contexts in which this is mitigated however–such as projecting the score or knowing it well–but these situations are often overlaid on the basic premise that discernibility is not necessary to the realisation or experience of the work: they might represent added value. Viewers may also not know how to decode any scores made available to them, rendering the impact of their visibility a moot point.

Some pieces, however, present perfect information such that all players know what is happening, and what has happened, and that these events may also be evident to the audience. In these situations, revealing the inner working of a piece offers up other ways to engage with the experience, potentially creating an affinity between players in their understanding of each other's choices, and empathy for the players from the audience. While this can in some cases remove the mystery of performance, it might also allow the experience to be more communal.

In this paper I consider the ways in which perfect information can be managed to create these correspondences in meaningful ways for participants and observers in scored compositions, focusing on open scores where decision-making processes are inherent to the mode of performance. My interest here is in creating a self-evident experience for observers, such that being welcomed inside the often-closed world of performance might shed light on the way we think and work together in groups.

2. INFORMATION

2.1 Perfect and Imperfect Information in Games

Our experience of games as observers is related to our understanding of what is happening. In games, the notion of perfect and imperfect information is crucial to our understanding of how processes operate, and correspondingly how games are played. The amount of extant knowledge we have regulates the choices that might be available to us, be they rational or not. In a game with perfect information, observers can also see what the players see and are able to speculate internally and strategise as a result. Such games flatten the differentiation between players and observers so that perceptible engagements with a game's components–moving a chess piece for example–are known to all. Perfect games may not explicitly reveal players' internal processing of game states, but the tangible results of their actions based on universally available information are. The availability of perfect information is referred to as an *extensive-form game* in game theory, where a game such as chess can be represented by a branching structure, the game tree, which outlines possible moves as a consequence of previous moves. We know, at least in theory, all possible moves from a given position. Observing a perfect information extensive-form game therefore allows us to empathise with the players by shadowing their decisions in a non-participatory manner. We can ponder what we would do next.

In imperfect information games, we may still receive gratification as observers by the reveal of hidden information by a player: the laying of a surprisingly effective card, the fortuitous dice roll, the opportune use of a power-up. Here too, while decision-making is less explicit as the field in which decisions are made is partly obscured, knowing the possibilities that might occur can open up similarly empathic responses predicated instead on the element of surprise. Our knowledge of the process, the system of interactions which unfold before us, may be sufficient to draw us in at a level where our understanding engages us in the strategy of players and their adaptability to circumstance. This is equally the case whether watching a cricket match or a livestreamed video game.

Elias, Garfield and Gutschera [2] refer to such perfect information as *public*, where players know things about the game, and imperfect information as *hidden*, where there are things they do not know. This may or may not be different for observers (e.g. peering over the shoulder of someone playing cards). They note the relationship between hidden information and spectation, suggesting that 'the more information is hidden, the worse the game's spectation' and that in hidden settings 'if the problem can be solved in some reasonable way. . . the audience may find the revelation of secrets quite appealing, leading to good spectation'. [2] In order to do this they suggest emphasising directional heuristics to 'make it easier for viewers to understand the state the game is in' and to let them say '"I would have done that differently!" and thus become more involved in the game.' [2] While this is framed within a gaming context, the idea of public and hidden information is relevant to spectation in music performance too.

2.2 Public and Hidden Processes in Music

Our experience of music as listener-observers is also related to our understanding of what is happening. In process music if the necessary information is made available to us, we have the potential to understand player choices and empathise with them. There is some commonality here with Steve Reich setting out the case for audible processes in music in his 1969 manifesto 'Music as a Gradual Process' [3]. He notes his wish 'to be able to hear the process happening through the sounding music' in such a way that we experience 'a compositional process and a sounding music that are one and the same thing'. While there is some ontological dissonance in this statement, the aim of revealing information that might normally be hidden as a way to shift attention towards a different kind of understanding is significant. Crucially Reich advocates doing this through the sounding music rather than through a separate explanation, such as a programme note or scholarly text.

This attitude is mediated through a compositional practice that focuses on gradual processes that unfold as we listen, rather than those buried within the compositional workings and seemingly obscured or privileged. But there are other ways to reveal aspects of these workings in the way a composition is experienced, such as listening with a score, whether this is through following it privately, viewing a score follower video where the audio and notation are synchronised, or by projecting the score for an audience. This can go some way towards the kind of revelation Reich advocates, but it is contingent on understanding the notation and being able to connect that in an analytical way with what is heard. It can be a specialist skill, and does not necessarily reveal information to all users.

It is important, therefore, to note the audience for this awareness of process. Later in the manifesto Reich comments that 'I don't know any secrets of the structure that you can't hear. We all listen to the process together since it is quite audible'.[3] This contrasts with many experiences of scored compositions, where the musicians realising a piece are permitted to glimpse aspects of its workings. As listeners we might recognise formal devices and technical models through exposure, but the contract is not with us directly.

But making processes audible does not necessarily equate with making the score public, although this can help. Rather it is about making available ways to understand what is happening in a piece more holistically. Sandeep Bhagwati suggests five different *conveyance modes* utilised in audio scores as a way to consider the best way to frame a scorer's intentions, such that 'each score type will need a different set of conveyance modes and will weigh their importance differently.' [4, p.26] These modes isolate instructional aspects of audio scores which may be useful in considering ways to share what is happening with observers. In particular his second category, instruction mode cues, contains four types of instructions:

- *Musical* instructions 'provoke musical structures that concern only the musician receiving the instruction' and might include commands to start or stop an activity, to play something previously specified such as a memorised pattern, or to play in a particular way.

- *Interactional* instructions 'concern the musical relations between two or more musicians', such as being instructed to accompany another player or emphasise a pitch heard in the soundscape.

- *Para-musical* instructions 'direct the performers to enact non-sonic behaviours' such as moving or other physical responses to the context.

- *Indexical* instructions 'point to, explain, and set up other conveyance modes' such as instructing one player to imitate another. [4, pp. 26–27]

If these instructions are made available to observers as well as participants following the instructions, the possibility of a self-evident music might result. In order to 'hear the process happening through the sounding music' however, enough of what is happening needs to be perceptible for observers. Self-evident processes suggest the possibility of close engagement with the operation of a piece, with the potential for empathic and communal experiences to be had by both participants and observers. If we are to consider what might be gained by presenting scores as systems of perfect information, we need to know what strategies and processes are available.

3. SELF-EVIDENT MUSIC

There seem to be relatively few pieces where the audience is made fully aware of the process as they observe it unfold. Composers use different modes to make this apparent to the audience, and these might form a partial framework for considering how such approaches could work. I suggest three intersecting ways this might happen: demonstrating, explaining and showing.

3.1 Demonstrating

The actions of the players may demonstrate non-verbally what is happening in the piece. The players might be tasked with an activity which they do in front of an audience. In undertaking this they do not explicitly explain what they are doing but their discussion and/or actions make this evident. In contrast to the notion of showing, demonstrations make processes self-evident through actions rather than presenting the score itself.

In Peter Ablinger's *Wachstum und Massenmord* (2009–10), a string quartet rehearse some unseen musical material for the first time in front of an audience. Ablinger notes that "The Rehearsal is the Piece. [...] The performers have not seen the score (except this foreword) ahead of time, and will receive it only directly before the performance, or, the scores are waiting for them on their music-stands." [5] Here the activity of the quartet references the rehearsal situation as a staged process of production. If we recognise and acknowledge this, aspects of the piece are traceable. The audience is not explicitly told what is happening, but it might be deemed self-evident that we are observing a kind of performative rehearsal.

Visible and audible cueing is another effective way of making player choices evident through demonstration. Depending on the nature of the instruction, interpersonal cueing can show aspects of each player's thinking, and the way such choices impact on the resultant music. I have explored this in my piece *it is the behaviour that a system tends towards and encourages that needs to be understood* (2021), where players cue material verbally, and then play it, creating connections with what other players are doing through explicit instruction or careful listening. The score pages comprise a series of 100 numbered cues, each of which indicate the starting point for the material chosen to be played. Other cues modify the way the material is played, including instructions to pause, change speed, register or direction of movement, and create looped sections. Players can give cues themselves and respond to those of others in order to investigate how concerted activity might emerge, co-ordinating and disrupting the structure of the group as a whole. The power structures within the group are playfully revealed, as are aspects of the performed personalities of the participants.

Other kinds of physical demonstration are possible, such as in Michael Baldwin's *a kind of nostalgia* (2014) where a guitarist playing a memorised piece of standard repertoire is controlled by mirroring the movements of another performer sitting opposite them. The other performer makes a series of fluid movements while holding a guitar of their own, and the guitarist must copy them, thereby rendering much of the playing impossible through the imposed contortions. No explanation is given, and the gradual exaggeration of movements that might occur leads observers through the process. It is of course questionable whether these demonstrations do reveal their workings to observers or whether it is just an aspiration: either way, the staged attempt is significant.

It should also be noted that the three categories – demonstrating, explaining and showing – intersect to form hybridised presentations. For example, in Robert Luzar's *Demonstrations* series (2016–20) videos or live performances "show audiences step-by-step actions on how to do – and possibly change – certain actions with their bodies, common materials and spaces." [6] These demonstrations are also partially explained by text captions or spoken commentary but often destabilise what is being observed.

3.2 Explaining

Pieces which involve demonstration can be allusive, whereas pieces which use explanation tend to reveal what is happening more explicitly using words. This may be through one or more of the players, or the composer, acting as a narrator and explaining what is happening to the audience. There are many examples of this, including pieces such as Tom Johnson's *Failing, a very difficult piece for double bass* (1975) and *Naryana's Cows* (1989) in which the narrator explains the process as it unfolds, Johannes Kreidler's *Fremdarbeit* (2009) where the composer explains the alleged outsourcing of the composition's making, or Matthew Shlomowitz's *Lecture about Bad Music* (2015) which playfully explores our habituation and relationship with music as listeners. All of these pieces involve spoken performative explanation of elements of the composition, such that these might be foregrounded as the listening focus. This could be a genuine attempt to clarify and make explicit what we observe, or it might distract from or obscure other aspects of the work, depending on what we believe.

In a similar way, pre-recorded voices can explain to the players what they must do. This explanation, and the actions that result, are apparent to the audience. Alvin Lucier's *I am sitting in a room* (1970) is a useful point of reference for this approach, with the score, and its realisation, turning this into a piece which describes its own making, enabling listeners to track the progress over time as the audio gradually flattens out. A more performative example is Louis d'Heudieres' *for* ____ *on* ____ (2015) which

> explores the idea of learning and rehearsing as performance. A group of performers hear a field recording for the first time. A voice explains that they should use it as a score, and have a limited time to learn it on their instruments. The audience is with them as they negotiate this task. [7]

The result is the players' attempt to do this within the temporal constraints set by the pre-recorded track. The audience follows the players' decision-making and can both think through their own approach and evaluate the players' choices as a result. Similarly in Stephen Crowe's *Tenvelopes* (2011–), horn player Samuel Stoll opens a series of ten envelopes, each containing a set of instructions for mostly uncomfortable activities such as licking the floor to partial undressing, guided by the (sometimes distorted) pre-recorded spoken voice of the composer. Although we do not know what is written on the cards, the framing of the wider piece makes it clear what is happening, and the enjoyment is in part from the comic timing and discomfort of the realisation of what Stoll is being asked to do.

Although I am focusing here on performative explanations which take place during the piece, it should be noted that external explanations such as programme notes, pre-performance talks, academic papers and interviews can contribute to developing an understanding of what happens in a piece. Some work is more ambiguous though, such as La Monte Young's *Composition 1960 #3* and *#4*, both of which require the activity of the piece to be announced to the audience.

3.3 Showing

Perhaps the simplest way to make what is happening evident is to show the score to observers, thereby removing much of the hidden information it contains. By doing this, it opens up aspects of the piece that may aid comprehension for some observers, although any understanding depends on being able to decode what is shown, whether this is stave-centred notation, tablature, static or moving images. For stave-centred notations, online score follower presentations are a clear way to connect what the notation shows to the audible musical responses from players, but this does require familiarity with these forms of notation. In a different way, animated notation can remove some of these barriers, such as in Ryan Ross Smith's *Study No.9* (2012) where players play one of their three sounds every time a moving cursor crosses an event node. Smith states in the performance instructions that "If possible, the Animated Score should be projected for both the performers and audience to see." [8] A common performance practice with animated notation is to project the score in this way, which, as David Kim-Boyle notes, "encourages an engagement with procedural relationships as they temporally unfold in the score and are musically sounded in the performance space." [9, p.48]

In some of the examples above, aural scores which explain or demonstrate what is happening intersect with the notion of showing in that observers have access to some of the same information as the players. Verbal explanation, in particular, may be a trace of the primary score, which sets out the process for what the players do and what observers experience: this is related, but different to showing the score itself.

3.4 Managing Information

Through the examples above I suggest three modes of delivering public information. The success of these approaches is contingent on the way such information is framed and managed for observers, such that it is understandable and meaningful. The factors which affect comprehension can also highlight the attitude towards revealing information encoded in a piece: some aspects might be hidden, deliberately made ambiguous, accidentally obscured, or revealed in absolute detail. The mode of delivering information and the method used to do it are both fundamental to this process, and comprise a set of broad parametric considerations that impact on the result. Four parameters are key to an understanding of what we perceive. The *speed* at which information is delivered has a direct bearing on how cues are understood. The *sequentiality* of information affects how we process cues in time with all the non-linear complexity of perception, memory, and expectation. The *density* of information regulates the audibility of separate cue elements and how we might differentiate them. The *simultaneity* of information affects how we manage parallel streams of cues and determines where our focus is placed.

Together these four parameters affect the cognitive load inherent in processing the information. In particular Sweller et al. note the difference in cognitive load between "the intrinsic nature of the material (intrinsic cognitive load) and the manner in which the material is presented (extraneous cognitive load)". [10, p. 57] Extraneous cognitive load is of particular relevance to comprehension in revealing the processes that are active in a piece. For example, in my piece *you are required to split your attention between multiple sources of information* (2018), multiple streams of audio cues articulated by six different artificial voices are occasionally densely layered, simultaneous and delivered at high speed making it difficult to isolate specific instructions. The extraneous cognitive load is high at these points, although it is introduced earlier in the piece in a more relaxed manner to aid comprehension and learning of the system by listener-observers. In all three modes of delivering public information, the cognitive load placed on observers, in addition to the players, has a direct bearing on how perceivable processes in the music are.

4. CONCLUSION

In the introduction, I asked what might be gained by presenting scores as systems of perfect or public information. Perhaps by revealing aspects of a piece which are normally unavailable through demonstrating, explaining and showing, the opening up of the experience may be seen as more welcoming for observers. Reducing the reliance on privileged knowledge, whether cultural or technical, has the potential to reduce barriers for engaging audiences. In some circumstances, this might suggest a kind of participatory engagement, one shared by observers of games or sports where thinking through strategies for possible moves or 'kicking every ball' reduces the space between them and more active participants. This sense of involvement-through-understanding can produce a feeling of empathy for participants from observers, especially where an activity has a tangible relation to everyday life and things that are already within our experience. This is not a suggestion that everything should be revealed, or that situations employing hidden or imperfect information are somehow lacking. Indeed there may be a concern that this kind of revelation removes the ineffable in music, despite Reich's note that "there are still enough mysteries to satisfy all." [3] But such self-evident processes in music can suggest the possibility of close engagement with the operation of a piece, with the potential for empathic and communal experiences to be had by both participants and observers.

5. REFERENCES

[1] K. S. Tekinbaş and E. Zimmerman, *Rules of play: game design fundamentals*. Cambridge, MA: MIT Press, 2003.

[2] G. S. Elias, R. Garfield, and K. R. Gutschera, *Characteristics of games*. Cambridge, MA: MIT Press, 2012.

[3] S. Reich, *Writings on Music, 1965–2000*. Oxford University Press, 2002.

[4] S. Bhagwati, "Elaborate Aaudio scores: concepts, affordances and tools," in *Proceedings of the 4th International Conference on Technologies for Music Notation and Representation*, 2018.

[5] P. Ablinger. "Wachstum und Massenmord," *Peter Ablinger*. Accessed: 27 October 2023. [Online]. Available: https://ablinger.mur.at/scores.

[6] R. Luzar, "Demonstrations," *Robert Luzar*. Accessed: 27 October 2023. [Online]. Available: https://www.robertluzar.com/demonstrations.

[7] L. d'Heudieres, "*for* _____ *on* _____" Louis d'Heudieres. Accessed: 27 October 2023. [Online]. Available: https://louisdheudieres.com/.

[8] R. Ross Smith, "Study No.9," *Ryan Ross Smith*. Accessed: 27 October 2023. [Online]. Available: https://ryanrosssmith.com/study9.html.

[9] D. Kim-Boyle, "Reframing the listening experience through the projected score," *Tempo*, vol. 72, no. 284, pp. 37–50, 2018. https://doi.org/10.1017/S0040298217001243.

[10] J. Sweller, P. L. Ayres, and S. Kalyuga, *Cognitive load theory*. New York: Springer, 2011.

EDGARD VARÈSE'S *POÈME ÉLECTRONIQUE*. AN ANALYSIS OF RECENTLY DISCOVERED SOURCES

Leo Izzo
independent researcher
izzoleo@gmail.com

ABSTRACT

The article investigates the genesis of *Poème électronique* by Edgard Varèse, analyzing several previously unknown or overlooked sources, which have been recently identified in the Varèse collection at the Paul Sacher Stiftung (Basel). The article examines two types of sources: Varèse's musical manuscripts and the technical plan for the sound spatialization of the entire piece, made by Philips engineers. While Varèse's sketches and drafts for *Poème électronique* had been neglected for years, the complete spatialization plan was once thought to be lost but has been recently rediscovered. These new sources reveal a largely unknown, long and complex process of creation, from the first music sketches to the final performance inside the Philips pavilion, a landmark in architecture designed by Le Corbusier for Brussels Expo in 1958. The article focuses on a short section of the piece, marked by Varèse as section F, that serves as a case study. Varèse composed this section by merging diagrams for electronic sounds, and conventional scores, such as fragments of his unpublished composition *Étude pour Espace* (1947). The spatialization plan for section F, coherent with Varèse's music manuscripts, suggests the vertical movements of drops falling from the ceiling. In future developments of this research, the composer's sketches and the project plan for the sound spatialization, both largely overlooked, could offer fresh insights for a deeper comprehension of the composition process and performance practice of this seminal work.

1. INTRODUCTION

In the tumultuous season of electronic music in the late 1950s, Edgard Varèse's *Poème électronique (PÉ)* stands as a striking exception that does not fit into the dominant trends of that period, namely Parisian *musique concrète* on one hand and the production of the Cologne studio on the other. Like his other compositions, Varèse did not provide a key for interpreting *PÉ*, nor did he attempt to explain his compositional techniques through theoretical writings. From a historiographical perspective, even more than fifty years after its creation, this work continues to pose challenges when trying to contextualize it within the electronic music production of the late 1950s. One of the reasons lies in the presumed lack of written sources relating to this work.

Varèse composed *PÉ* for a multimedia performance at the Philips Pavilion during Expo 58, the Brussels World's Fair of 1958. The visual aspect of the performance was overseen by architect Le Corbusier who also designed the futuristic structure of the building, with the crucial assistance of his collaborator Iannis Xenakis. From September 1957 to April 1958 Varèse worked in a dedicated Philips facility, the garage at Philips' Strijp III complex in Eindhoven, near Philips Research Laboratories. As Kees Tazelaar described in detail [1], Varèse had access to a state-of-the-art laboratory for electronic music production, with the assistance of expert Philips technicians such as Willem Tak, Jan de Bruyn, and Anton Buczynski. Varèse recorded *PÉ* on three mono tapes that were played simultaneously on three tape recorders. A lead-in tape, which contained sync and count, enabled the tapes to be approximately synchronized. When the recordings were completed, a fourth stereo tape was added to the three master tapes, in order to add reverberation. The four tapes were then mixed into a special three-channel perforated 35mm tape used in the film industry. The perforations assured synchronization with the four projected films and with an additional perforated tape containing the control signals for the fully automated performance.

In total, 325 loudspeakers were mounted on the walls, and an additional 25 speakers for low frequencies were positioned around the perimeter. This setup allowed the sound to resonate from all directions, revolving around the spectators or hovering above them from a height of approximately twenty meters, at the highest point of the pavilion. Today, it is no longer possible to experience this work in its original form. The pavilion was demolished shortly after the closure of the Expo, and with it, the only device capable of reproducing this multimedia spectacle was lost. Since the demolition of the pavilion, *PÉ* has primarily been recognized as an independent musical piece. This recognition is largely due to a version created for the album *Music of Edgar Varèse* (Columbia Records, 1960, MS 6146), which simplifies the original multi-directional sound spatialization into a stereophonic recording. Studies on the sources of *PÉ* have mainly focused on the tapes [2], the setup of the audiovisual system [3], [4], and on Le Corbusier's scenario for light projection [5]. However, there are two aspects of this work that remain almost completely unknown: Varesè's compositional process and the sound spatialization during the performance.

This article focuses on these two aspects, referring to written sources that are still almost entirely overlooked: the musical manuscripts that Varèse wrote during the composition of the piece and the project for the sound spatialization of the entire piece. The central purpose of this publication is to outline a philological method for examining the sources relevant to *PÉ*, encompassing the early sketches to the ultimate plan for sound projection. Since the analysis of the entire *PÉ* is beyond the scope of this paper, the explanation will focus on a short section of the piece.

2. VARÈSE'S COMPOSITIONAL PROCESS THROUGH THE ANALYSIS OF THE SKETCHES

Varèse's manuscripts became accessible to scholars in 2003. These were part of the Edgard Varèse Collection, which was previously curated by his pupil, the composer Chou Wen-chung. The collection is now housed in the Paul Sacher Foundation in Basel, henceforth referred to as EVC-PSS. Many sketches related to *PÉ* were overlooked for years as they were dispersed among the composer's numerous unidentified documents. Their extreme heterogeneity has certainly constituted an obstacle to identification. Indeed, this collection of documents features a wide variety of notation styles. It includes multicolored plans for the electronic montage, symbolically written annotations, diagrams on graph paper, draft scores for electronic pitches in standard staff notation, sketches of short rhythmic elements, and the use of Cartesian axes to notate the frequency or the envelope of sounds. The diverse nature of the documents partially reflects the characteristics of the work itself. Compared to most electronic music of that era, it stands out due to the extensive variety of sound materials used.

However, the diverseness of these manuscripts also unveils the various work methods that Varèse adopted in the preparation of *PÉ*. Through a systematic comparison of different sources, primarily preserved at the Paul Sacher Stiftung (PSS, Basel), I could identify more than fifty manuscripts related to *PÉ*.[1] Among these, I recognized a small group of pre-definitive manuscripts, or 'drafts', which Varèse likely used as a reference for the final realization and assembly of the tape. In two sections of the piece (from 3:38 to 4:39, and from 6:44 to 7:06), electronic sounds are almost entirely absent. These sections feature two 'acoustic' episodes: the first one for contralto voice, male choir, percussion, and piano, and the second one for soprano and male choir. To record these two episodes, Varèse prepared a specific score by reworking a piece from 1947, *Étude pour Espace* (*Et.58*, EVC-PSS). This score represented the final attempt to bring an ambitious project to fruition. Since the mid-1920s, Varèse had been chasing the concept of a grand multimedia opera, which he referred to by various titles over time, including *The One-all-alone*, *The Astronome*, *Sirius*, and *Espace* [6], [7].

As a result, the sketches for *PÉ* include not only electroacoustic diagrams but also some pages in standard notation intended for a small ensemble. Therefore, the central core of the sketches consists of nine electroacoustic diagrams, in addition to the two score excerpts from *Et.58* (Figure 1).

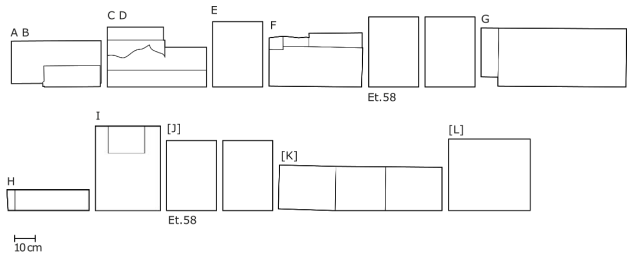


Figure 1. Sequence of *PÉ* final drafts.

These drafts correspond to various sections of the piece, each lasting from fourteen seconds to a minute and thirty seconds. A comparison with the tape reveals a significant correlation between the notated parts and the recorded sounds. The sequence of the manuscripts is confirmed by a final mark: a capital letter in red pastel, likely added to almost every draft upon completion. Some manuscripts encompass two sections and, consequently, bear two alphabetical markings. Without a comprehensive structural plan for the entire piece, these red letters, spanning from A to I, serve as the sole indicators of the sequence of electroacoustic plans, and thus, the arrangement of the sections in the piece.

The eighth and ninth manuscripts lack the red mark. They were included in the group of final drafts by deduction, as they are the only existing detailed plans corresponding to the final part of *PÉ*.[2] By arranging the sketches and the fragments derived from *Et.58* in the correct sequence, one can recognize a written plan for almost the entire piece.

[1] This identification work started in 2008 with a scholarship from the Paul Sacher Foundation, and continued in the following years in other archives, including the Philips Company Archive of Eindhoven and the Fondation Le Corbusier of Paris. Important archival materials relating to *PÉ* are also held in other institutions: Louis Christiaan Kalff materials of the Getty Research Institute (Los Angeles), the Stanford University Libraries; the Dutch Film museum EYE (Amsterdam), the Institute of Sonology (Den Haag).

[2] Many evidences suggest that during an advanced stage of composition Varèse decided to reassemble the already recorded sections of the tape. In order to establish a definitive order in the structure, he added the red letters to the existing drafts. Presumably after this reworking, he begun to write the last three final drafts, without the need of other alphabetical marks. I explained in more details this stage of composition in the article "La genesi compositiva di 'Poème électronique' di Edgard Varèse" [1].

As Anne Shreffler recalled, “Varèse’s inexperience with tape composition has always been known in electronic music circles” [8]. This assumption influenced also the study of his only electronic work and, as a result, scholarship devoted little interest to his manuscripts for *PÉ*. One remarkable case is his final draft for section C and D. This manuscript has been known since 1958, when Jean Petit reproduced it in the volume *Le Poème électronique: Le Corbusier* [9]. Later on, it was reproduced in several publications: it appears on the cover of *Edgard Varèse und das ‘Poème électronique’: eine Dokumentation*, edited by D. A. Nanz [10] and in Peter Wever’s *Inside Le Corbusier's Philips Pavilion: a multimedial space at the 1958 Brussels World's Fair* [11]. However, its content and the exact relationship with the tape remained unclear, due to the difficulty in understanding it without the other missing pieces, until my research in 2015 [12]. Among the studies devoted to music sketches, the dissertation by Roberta D. Lukes is noteworthy. However, in 1996, she had access only to a few manuscripts held in the Philips Company Archive in Eindhoven [13].

The abundance of Varèse’s sketches and drafts found in PSS attests to his efforts during the composition process of *PÉ*. At the age of 75, with no previous experience in electronic music production, Varèse was attempting to handle new instruments and confront new technical and musical challenges.

While Varèse’s music manuscripts display detailed information about sound events, they contain almost no annotations about sound spatialization. Varèse noted the spatial possibilities of sound projection in his notebook immediately upon his arrival in Eindhoven, but his drafts for *PÉ* contain very few annotations for sound distribution. The only written plan for sound spatialization was prepared by Philips technicians, likely after the recording on three tapes was completed. Moreover, Willem Tak asserted that Varèse did not participate in the planning of sound spatialization within the Pavilion, and that the decisions were made by Philips technicians [14].

3. *PÉ*, SECTION F

In order to clarify Varèse’s compositional process during the creation of *PÉ*, we can use section F (from 3:25 to 4:39 of the tape) as an exemplary case study. This section is composed of a tape edit of two distinct preparatory sound sources: a sequence of percussive electronic sounds that were planned out on two technical diagrams, which we will refer to as ‘a’ (as shown in Figure 2) and ‘b’, and the first episode derived from *Et.58*, which features a contralto voice, choir, and percussion. Section F begins with these two elements alternating for approximately 50 seconds (from 3:25 to 4:15), followed by the recording of percussion and choir until the end of the section (4:39).

The tape editing and the related sources for the first 50 seconds of section F is shown schematically in Figure 3.

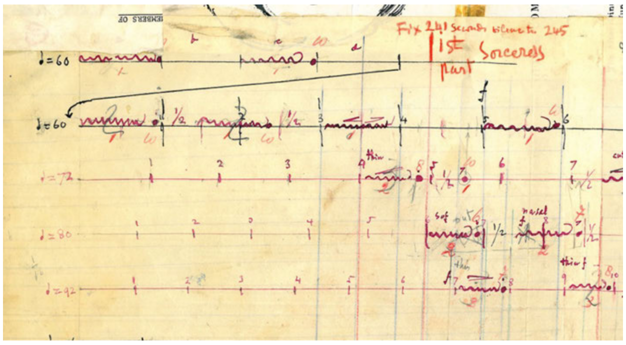


Figure 2. Diagram *a,* draft for section F of *PÉ*, excerpt. ECV-PSS. Courtesy of PSS.

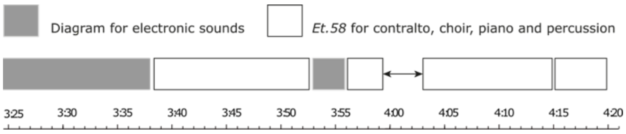


Figure 3. Sources and tape editing in section F (3:25–4:20).

Within Varèse’s collection in Basel, there is no evidence of a comprehensive editing plan for *PÉ*. However, Varèse annotated on the music manuscripts the junctions between the electronic percussive sounds and the vocal part. In the revision of *Et.58*, the initial entry of the contralto is preceded by the annotation “Perc[ussion] – introducing” (Figure 4).

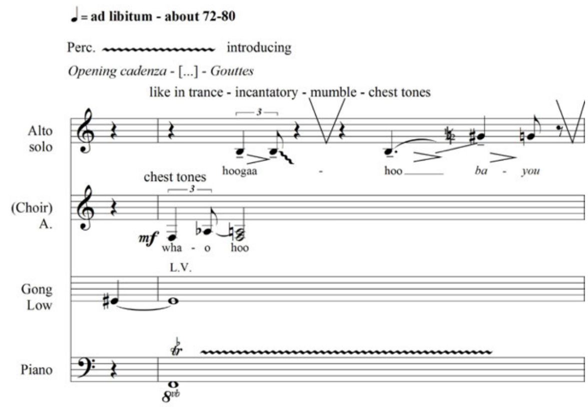


Figure 4. *Et.58* (excerpt), diplomatic transcription (EVC-PSS)

An additional annotation at the bottom of the same page specifies: “sporadic p[e]rc[ussion] electronic”. These notes serve presumably as a reminder for Varèse himself: an annotation for a passage to be developed later. In the same document, the term ‘gouttes’ (French for water drops) in pencil describes the overall effect of this percussive “opening cadenza”. The electronic diagram created for section F can thus be interpreted as the realization, in the form of a detailed draft, of the embryonic idea of the “sporadic p[e]rc[ussion] electronic” initially envisaged in *Et.58*.

Conversely, the electronic diagram contains references to the *Et.58* score, inserted as notes for the editing along the upper edge of the diagram. With the phrase "1st Sorceress part" written in red crayon, Varèse notes where to insert the contralto recording. For these reasons, both manuscripts (the electronic diagram and the *Et.58* score) can be considered integral elements of Section F, despite their distinct appearance and sonic result. Like the other final drafts, this electronic diagram also has an alphabetical marking. However, the 'F' marked in red crayon is not clearly visible as it is obscured by a strip of paper (this can be slightly discerned in the upper left corner of Figure 2).

4. ELECTRONIC DIAGRAM FOR SECTION F

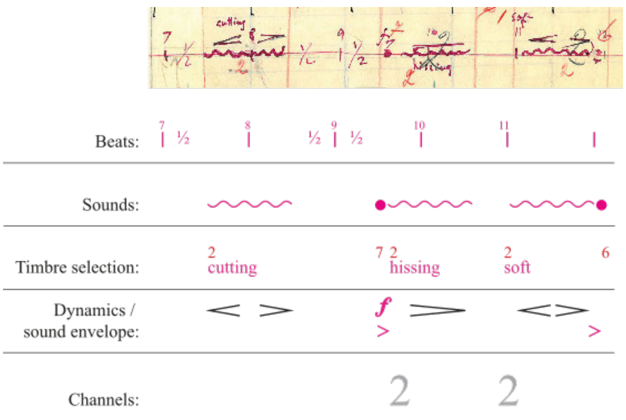


Figure 5. Notation layers in diagram *a* for section F.

In this diagram, Varèse notated four distinct aspects of this electronic episode. The first aspect pertains to the superimposition of simultaneous rhythmic aggregates based on 'unrelated' metronomic tempos, a recurring element in the sketches of *PÉ*. The overlay of percussive entries is structured in accordance with four concurrent metronome tempos, corresponding to 60, 72, 80, and 92 bpm. This polytemporal grid served as a foundation for the preparation of the draft. These four lines were used to notate the individual sounds whose duration, generally, coincides with the tempo stipulated by the metronome mark.[3]

The second aspect pertains to the timbre and morphology of each sound, reflecting the development of sound over time. With a few exceptions, each 'drop' is composed of a pair of sounds: it begins with a sustained noise, notated with a wavy horizontal line, and concludes with a short pulse, represented as a point. The episode involves a process of subtle timbral differentiation among the elements of the electronic percussive sounds.

To select the right sound source, Varèse added a red crayon number (ranging from 1 to 10) beneath each undulated line, likely referring to a reserve of materials at his disposal. Moreover, certain timbral variants are verbally described with adjectives such as "thin", "cutting", "hissing", "soft", and "nasal". Figure 5 reproduces a portion of the sketch relating to beats 7–10. The different layers of text, sometimes overlapping each other, are displayed separately in the graphic below: beat numbers and the duration of rests, the type of sound (sustained noise/pulse), the selection of timbres, the dynamics marks, and the recording channel. The third aspect concerns the distribution of sounds in the three channels (or, more precisely, mono tapes). Compared to the other manuscripts, the draft for section F contains particularly detailed information regarding the distribution of sounds in the three channels. In this draft too, as in other music manuscripts, there is no information about the distribution of sound inside the Philips pavilion. Nevertheless, the careful organization of the three channels seems to be preparatory work for the subsequent spatialization. The assignment of sounds to the three channels took place in two stages. First, in diagram *a* (Figure 2) Varèse assigned each element to a channel, drawing a large number in pencil. The sounds noted in the excerpt of Figure 5 are assigned, for example, to channel 2. Afterwards, to facilitate the technical realization of the three mono tapes, Varèse rewrote the entire episode in the new diagram *b*. To make the content of each mono tape clearer, he distributed the sounds in a new layout, grouping them according to the destination channel (Figure 6).

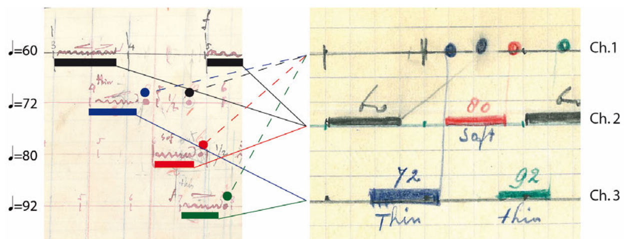


Figure 6. Distribution of sounds in diagram *a* and diagram *b*. Courtesy of PSS.

The graphic elaboration of Figure 6 illustrates the conversion process from diagram *a* (on the left) to diagram *b* (on the right). The prolonged noise sounds, which at first were wavy lines, now become horizontal bars and are assigned to channels 2 and 3. The pulses, on the other hand, are represented with colored points and converge in Channel 1.

[3] Several years after *PÉ*, Varèse continued to be attracted by the possibilities of managing complex rhythmic events through electronic means. Interviewed in 1965 about the potential of electronic instruments, Varèse said: "One of the most valuable possibilities that electronics has added to musical compositions, at least for me, is that of the possibility of metrically unrelated simultaneity. My music being based on the movement of unrelated sound masses, I have long felt the need and anticipated the effect of having them move simultaneously at different speeds" [15].

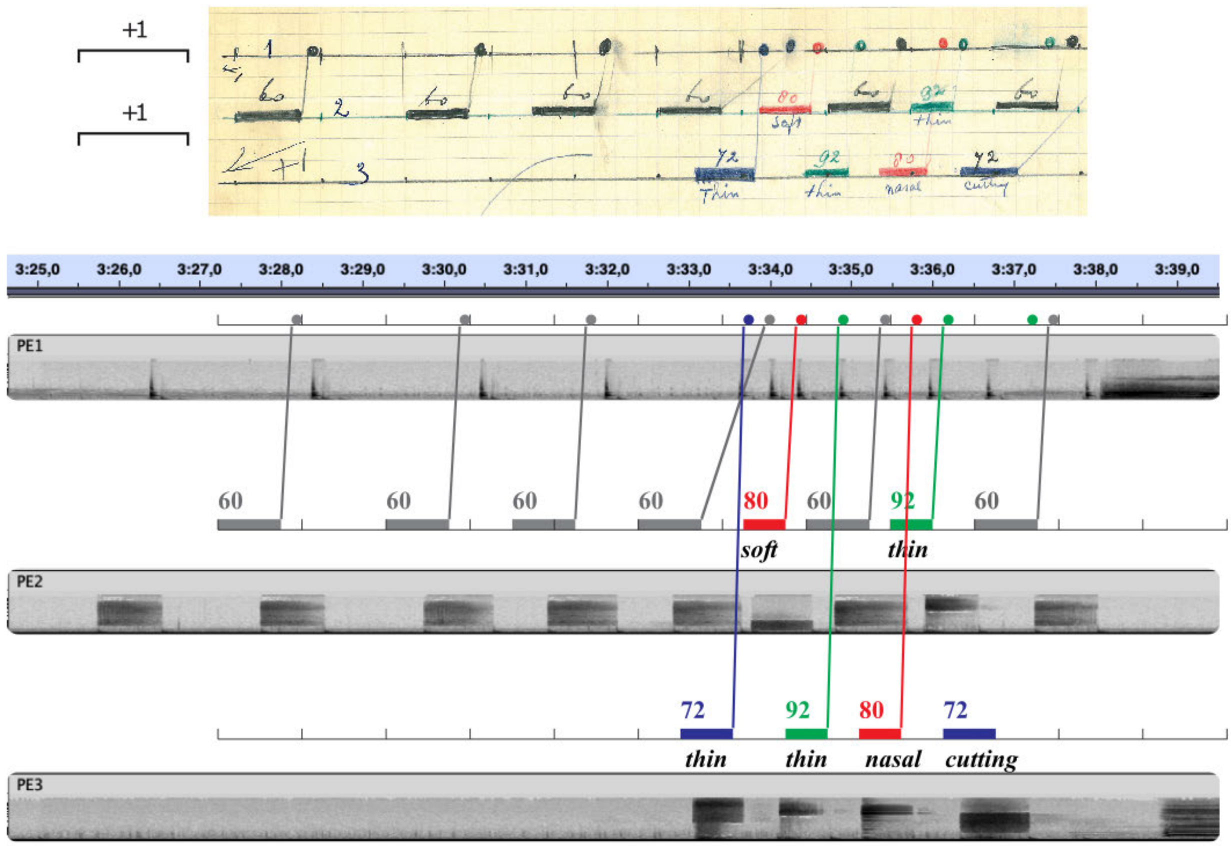


Figure 7. Correspondences between diagram *b* and the tape (from 3:25 to 3:39). Courtesy of PSS.

A color code differentiates the tempo: black for 60 bpm; red for 80 bpm; blue for 72 bpm; green for 92 bpm. In the two drafts, the content remains the same. However, the two graphic renditions highlight either the tempos (in diagram *a*) or the distribution of sounds among the three channels (in diagram *b*). Except for some slight temporal discrepancies, the events recorded on tape faithfully reflect the written traces. The scheme of Figure 7 displays a graphic combination of diagram *b* and the sonogram of the three mono tapes (from 3:25 to 3:39). The corresponding portion of the manuscript is shown above.

5. SOUND PROJECTION

The distribution of sound within the Philips Pavilion represents one of the most innovative and spectacular features of *PÉ*, yet it is also one of the least documented aspects. The volume *Poème électronique* edited by Jean Petit includes a diagram by Xenakis titled “Les routes du son” (Figure 8) [9]. This drawing provides a brief description of the distribution of the speakers within the Philips Pavilion. The trajectories of the sound are marked with the numbers I–IX, and the letter O, while the position of the speakers is marked with letters A–E, J, and U.

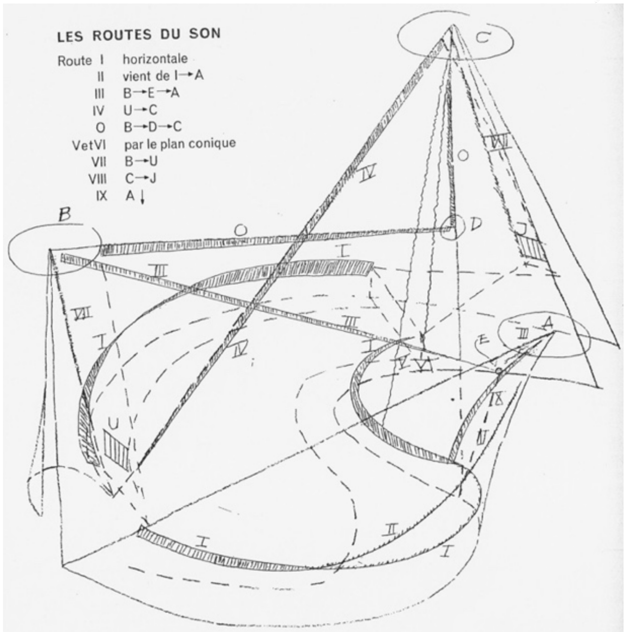


Figure 8. Iannis Xenakis, *Les routes du son*: diagram for the spatialization system in Philips pavilion

This drawing is accompanied by an explanatory legend. The speakers were arranged in two configurations: some groups, referred to as ‘clusters’, served as static sources, while others, referred to as ‘routes’, simulated the movement of sound along a line.

The clusters A, B, and C were located on the three peaks of the building (at 13, 18.5, and 20.5 meters respectively). The clusters J and U were positioned at the entrance and exit as antiphonal sources. Finally, other sources for the lowest sounds surrounded the audience at floor level along the perimeter of the building.

Kees Tazelaar [1] conducted extensive research on the sound projection system inside the Philips pavilion and almost entirely reconstructed the placement of the speakers on the walls.

However, until recently, it was not possible to determine the sequence of the different spatialization scenarios during the performance. For a long time, the only source explaining the sound distribution during the piece had been a diagram written by Philips technicians and published in the journal *Philips Technische Rundschau*, with the caption "Schematische Darstellung der registrierten Klänge" (Schematic representation of the registered sounds) [15]. This diagram illustrates how the control tape could synchronize tape recorders, lights, loudspeakers, and the film projector through a sequence of multiple signals. Unfortunately, it is only an excerpt related to a short portion of the piece: from 2:05 to 2:42. This control diagram has a timeline at the top and is divided into two horizontal stripes. As described by V. Lombardo and other authors, the upper stripe illustrates the contents of the control tape [3].

The sound impulses recorded on the control tape operated on the different devices through a system of filters and relays. The lower stripe schematizes the three audio channels. Each short episode in each channel is symbolized with a horizontal line that defines its duration. The most interesting data in this diagram relates to the spatialization. Every line (or tape segment) is associated with a letter or a number. These symbols refer mostly to Xenakis's drawing of Figure 8.

Outside of these thirty seconds, there were no other clues about the sound projection during the performance. However, a recently found source, originally held by Wen-chung's heirs and now in PSS, fills this lack of information. It consists of an approximately 10 meters long scroll paper. It is clearly a complete copy of the original technical diagram that was made by Philips technicians for the sound spatialization and later sent to Varèse. As already mentioned, without any other autograph document, it is difficult to establish who made the decision about sound spatialization. Nevertheless, this document is an invaluable source for understanding and reconstructing the performance of *PÉ*, including the sound projection, as it contains detailed information on the sound sources, the routes, and the timings of each sound event.

In general terms, the sound distribution reflects the articulation of the piece into sections. The diagram shows clearly that spatialization scenarios were used to enhance the contrast between subsequent sections. One recurring means is the opposition between a scenario with several sound sources and a rapidly changing movement of sound, and another scenario with static sound sources. These complementary principles are very clear in the spatialization of sections D and E. The beginning of section D, from 1:10 to 1:30, is particularly dense, with a fast edit of complex noises that differ in morphology, shape and timbre. For this part, the spatialization diagram lists eleven different sound sources (trajectories and speaker clusters) in only twenty seconds. One can imagine the audience listening to scattered noises, with their origins rapidly changing within the pavilion. On the contrary, section E, from 2:40 to 3:25, is characterized by sustained electronic tones. The corresponding sound distribution is limited to the three peaks of the building and to the perimeter, resulting in a more static effect surrounding the audience. Since the analysis of the sound distribution for the entire piece is beyond the scope of this paper, I will focus on section F only, to highlight the correlation between different sources: sketches, tapes, and the technical spatialization plan. The portion of the spatialization diagram related to section F reveals some interesting aspects. Each sound is labeled with a short description, mostly in Dutch, using three types of terms: 1) onomatopoeias and words unrelated to Varèse's sketches, such as "plop"; 2) descriptions of recorded sounds, such as "piano" or "koor" ('choir' in Dutch); 3) quotations from Varèse's own words, such as "sorc.[eress]" or "drup" ('drop' in Dutch, equivalent of 'goutte'). The information about spatialization reveals clearly the intention to simulate the falling of each electronic 'drop' from the ceiling to the floor of the pavilion, especially at the beginning of the section, as shown in Figure 9. The sustained noise sounds (channels 2 and 3) were distributed along *route* O, a line of speakers located above the audience, while the subsequent electronic pulse originated from the line of speakers labeled as 'L.t.', situated at the lower perimeter of the building. Here, 'L.t.' likely stands for 'Low tone'. The voice of the contralto singer was projected from the left perimeter of the building ('I l.[eft]'), while the choir sounds were projected from the perimeter ('L.t.') and from one of the peaks of the building ('A').

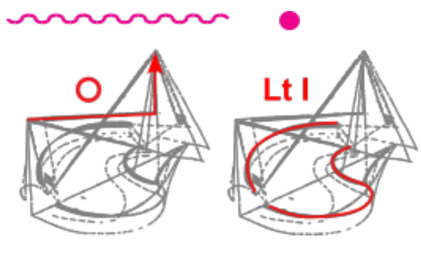


Figure 9. Spatialization of 'drops': correspondences between diagram *a* and Xenakis' *Les routes du son*.

Figure 10 illustrates the sound spatialization for section F. The scheme combines different sources: the spectrograms of the three mono tapes are aligned with the information about the sound projection of each track. This information is transcribed from the control diagram (refer to the horizontal lines: I, II, III). Several miniatures from Xenakis's drawing are displayed to emphasize the corresponding routes or clusters of speakers within the Pavilion.

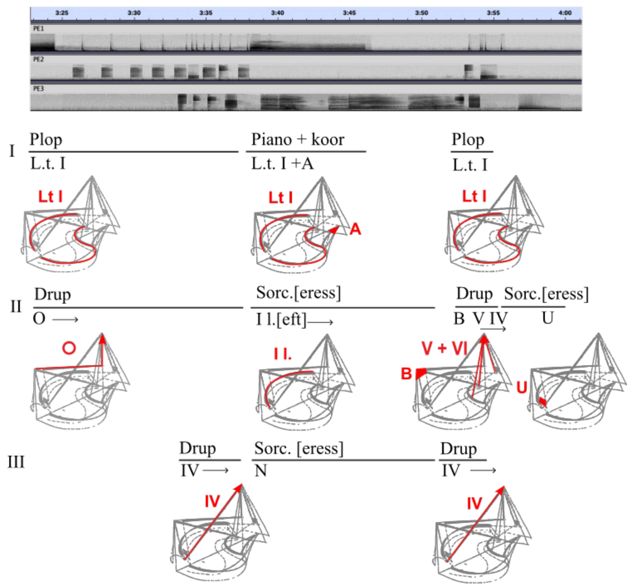


Figure 10. Sound spatialization in *PÉ* (3:25–4:00). Graphic elaboration from Control Diagram, Xenakis' *Les routes du son* and the sonograms of the tree mono tapes.

6. CONCLUSION

Varèse's sketches and drafts and the control diagram for spatialization made by Philips technicians represent a fundamental step forward in the comprehension of this seminal work in the history of electronic music. These sources reveal the internal structure of the composition, including the routing of sound through the complex sound diffusion system of the Philips pavilion. Previous studies on *PÉ* have followed mainly three directions: the structural analysis of the piece through listening to the audio recording [17]; the historical research on the technical system of sound distribution [1] and the architecture and light projection [11]. The discovery of new sources has ushered in a new phase in the study of Varèse's compositional thought. Moreover, these sources provide the opportunity to perform *PÉ* in its original form, grounded on philological research.

7. REFERENCES

[1] K. Tazelaar, *On the Threshold of Beauty. Philips and the Origins of Electronic Music in the Netherlands 1925–1965*, Rotterdam: V2_Publishing, 2013.

[2] L. Cossettini, "The performance edition of Poème électronique by Edgard Varèse", *Sounds, Voices and Codes from the Twentieth Century. The critical editing of music at Mirage*, ed. by. Luca Cossettini and Angelo Orcalli, Udine: Mirage, 2017, pp. 389–400.

[3] V. Lombardo et al., "A Virtual-Reality Reconstruction of 'Poème électronique' Based on Philological Research", *Computer Music Journal*, vol. 33, n. 2, 2009, pp. 24–47.

[4] J. de Heer and K. Tazelaar, *From Harmony to Chaos: Le Corbusier, Varèse, Xenakis and Le Poème Électronique*, Amsterdam: Uitgeverij Duizend & Een /1001 Publishers, 2017.

[5] M. Treib, *Space Calculated in Seconds: The Philips Pavilion, Le Corbusier, Edgard Varèse*, Princeton: Princeton University Press, 1996.

[6] M. MacDonald, *Varèse: Astronomer in Sound*, London: Kahn & Averill, 2003, pp. 216–241.

[7] A. Jostkleigrewe, "Reaching for the Stars: From 'The-one-all-alone' to 'Espace', in *Edgard Varèse: Composer, Sound Sculptor, Visionary*, ed. By F. Meyer e H. Zimmermann, Woodbridge: Boydell, 2006, pp. 211–219.

[8] A. Shreffler, "Varése and the Technological Sublime; or, How lonisation Went Nuclear", *Edgard Varèse: Composer, Sound Sculptor, Visionary*, ed. by Felix Meyer and Heidi Zimmermann, Woodbridge: Boydell & Brewer, 2006, pp. 290–297.

[9] *Le Poème électronique: Le Corbusier*, ed. by Jean Petit, Paris: Les éditions de minuit, 1958.

[10] *Edgard Varèse und das 'Poème électronique': eine Dokumentation*, ed. by D. A. Nanz, Basel: Paul Sacher Stiftung, 2006.

[11] P. Wever, *Inside Le Corbusier's Philips Pavilion. A multimedial space at the 1958 Brussels World's Fair*, Rotterdam, Nai010 Publishers, 2015.

[12] L. Izzo, "La genesi compositiva di 'Poème électronique' di Edgard Varèse", *Il Saggiatore musicale*, vol. 22, No. 1, 2015, pp. 61–96.

[13] R. D. Lukes, *The "Poème électronique" of Edgard Varèse*, Ph.D. Diss., Harvard University, 1996

[14] W. Tak, "Verwirklichung des 'Elektronischen Gedichtes' im Philips-Pavillon auf der Brüsseler Weltausstellung. B. Die Klangeffekte", *Philips Technische Rundschau* 3/4 (1958/59), pp. 75–77.

[15] V. Perlis and L. Van Cleve, *Composers' Voices from Ives to Ellington: An Oral History of American Music*, New Haven: Yale University Press, 2005, p. 110.

[16] S. L. De Bruin, "Verwirklichung des 'Elektronischen Gedichtes' im Philips Pavillion auf der Brüsseler Weltausstellung – C. Die Elektronische Steuerung", *Philips Technischer Rundschau*, n. 3/4, 1958/59, pp. 77–82.

[17] G. S. Kendall, "Juxtaposition and Non-motion: Varèse Bridges Early Modernism to Electroacoustic Music", *Organised Sound*, vol. 11, 2006, pp. 159–171.

CHALLENGING EPISTEMIC BIASES IN MUSICAL AI: A GUERRILLA APPROACH TO HUMAN-MACHINE COMPROVISATION BASED ON XENAKIS'S SKETCHES FOR *EVRYALI*

Pavlos Antoniadis
Department of Music Studies,
University of Ioannina
p.antoniadis@uoi.gr

ABSTRACT

The objective of this paper is to reflect on the affordances of sketches as interfaces for human and machine learning, by way of a case study based on Iannis Xenakis's *Evryali* (1973). First, I report on one-to-one mappings between the composer's original sketches and the symbolic notation intended for performance. Then, there is an outline of the sketches' deviations from the symbolic score and their potential to offer indispensable analytical insights for learning. The decoupling of sketch and score intensifies as performance multimodal data enters the framework of my analysis, allowing for the emergence of one-to-many mappings among those three distinct representation domains. This multiplicity of relations fuels the creation of gesture-controlled, augmented, and interactive tablatures, which are based on the sketches and incorporate graphic and multimodal elements to bypass conventional notation. Finally, I report on the use of tablatures as both preparation and performance tools in a human-machine comprovisation setting, involving a human trained to improvise on complex scores, an AI agent trained on a corpus of recordings, and a gesture-follower trained on the performance of sketches. As a postlude, the potential of one-to-many mappings for challenging established epistemic biases in musical AI is stressed. I capitalize on the unpredictability generated by the interplay between couplings and decouplings of different representation domains, affirming the transitory nature and inherent malleability of sketches.

1. XENAKIS AND GRAPHICS

Although the relationship of the architect, engineer, and composer Iannis Xenakis to graphic design might seem too intuitive to stress, some of his writings offer a more convoluted image.

The first element to point out is that sketches facilitated Xenakis's control over global formal properties that are not readily accessible in the symbolic notation of serial linear polyphony. His proposition of "a world of sound masses, vast groups of sound events, clouds, and galaxies governed by new characteristics such as density, degree of order, and rate of change" [1, p. 182] wouldn't have been possible without both probability theory and the visual means to implement it. The sketch is integral for the application of stochastic and probabilistic laws that were hitherto impossible.

Second, Xenakis dissociates the term "sketch" from its graphic implementation when he, for example, refers to symbolic music as a "logical and algebraic sketch of musical composition." [1, p. 155] This conceptual dissociation becomes clearer as Xenakis warns against the "fetish of the graphic symbol", whereby "the music is judged according to the beauty of the drawing." He does so in favour of a functional or algorithmic perception: "[...] graphical writing, whether it be symbolic, as in traditional notation, geometric, or numerical, should be no more than an image that is as faithful as possible to all the instructions the composer gives to the orchestra or to the machine." [1, p. 180]

A final word should be uttered in relation to Xenakis's theory of musical time and the distinction between *temporal*, *inside time* and *outside of time* materials [2]. The distinction indicates a dialectic between the graphic representation of instructions and the clash they generate between lived experience and fixed architectures on a "blank blackboard of time, on which symbols and relationships, architectures and abstract organisms are inscribed." [1, p. 192]. This dialectic of lived and abstract time might later prove indispensable for the leap from a score reproduction mode characteristic of high modernism [3] to a multimodal comprovisation mode based on sketches, or the leap from one-to-one towards one-to-many mappings.

In what follows, I opt for a bottom-up approach. I deterritorialize the idea of a sketch from the inside, showing how even the simplest mappings can produce a variable matrix of performance possibilities, before being further repurposed towards other goals, such as gesture following based on machine learning and Human-AI coadaptation based on audio and MIDI recordings. Artistic "guerrilla" practice complements a strict methodology of mappings and functionalities, allowing for a vital counterpart to Markov Models and Factor Oracles, the human(-in-the-loop)[1] [4] as "ghost in the machine" (after Gilbert Ryle).

2. ONE-TO-ONE MAPPINGS

2.1. Simple

A reproduction model of the musical score assumes a tight coupling between symbolic elements and their sketch counterparts. This will also be my own departure point, albeit with the intention of discovering decouplings, which render sketches indispensable.

Figure 1 presents a case of obvious one-to-one mapping of pitch and rhythm information. Bar 1 is annotated in red to show pitch correspondences between sketch and score, and bar 3 in blue for rhythm information respectively. One cell per semitone on the vertical axis and one cell per 16th note on the horizontal axis of the sketch allow for an accurate representation of pitches, bar lines, and attacks' positioning (or composite rhythm), but crucially not the duration of individual attacks. Indicating duration has required an extra annotation layer of blue lines for the 16th and orange lines for the 8th notes (bar 3).

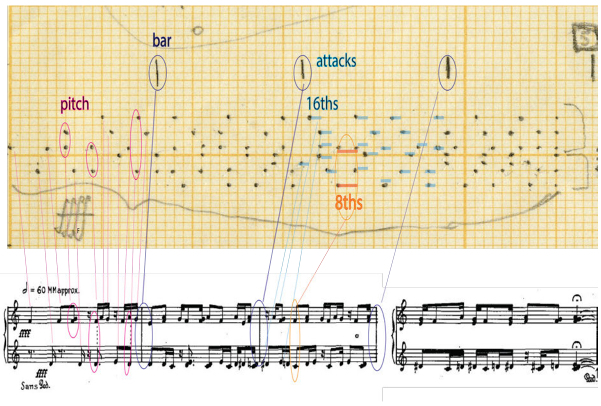


Figure 1. One-to-one mappings in the pitch and rhythm domains. Sketch reproduction of *Evryali*, bars 1–4, with kind permission by the Xenakis Archive – Mâkhi Xenakis

2.2. Complex

Figure 2 presents a more nuanced case, whereby pitch and rhythm information is complemented by information on texture and form that is not afforded by the symbolic score. First, a notion of "isodynamic lines polyphony" features in the sketch as lines connecting pitches bearing the same dynamic (texture B in pink frame). This texture is complemented by a distinct texture based on repeated notes (C in purple ellipse), which appears to be an embellishment rather than an interruption of the general form (D in green ellipse). Both features, the polyphony of dynamic lines and the hierarchy of the two textures, are hard if not impossible to discern in the pointillistic symbolic score. In that sense, the symbolic score presents lower affordances for performance, whereas the sketch presents higher-order information decoupled from the performers' score.

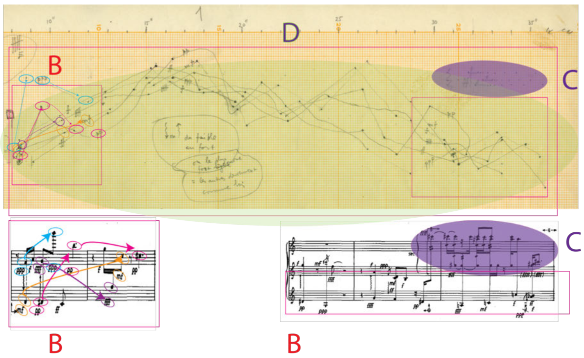


Figure 2. One-to-one mappings that reveal textures and formal properties not readily accessible in the symbolic score. Sketch reproduction of *Evryali*, bars 5–18, with kind permission by the Xenakis Archive – Mâkhi Xenakis

2.3. Global

In Figure 3, the complete annotated *Evryali* sketches have been assembled into a single representation, which provides a rapid understanding of the work's global form. This understanding is based on meticulous analysis of texture, as well as on the difficulty and occasionally impossibility of realizing the textures in performance, as a measure of the work's complexity. Different colours indicate different textural types. Each of the original composer's sketches contains two lines. Each textural frame is indexed through a verbal description, its correspondence to the published score by *Editions Salabert* (page and bar numbers) and the track number, indicating reference recordings.

According to this analysis, the articulation of the piece's form is governed by the following parameters:

a.) Alternation between blocks of points and linear arborescences [5] of varying complexity. For example: First theme texture (as in Figure 1) versus second theme texture, as in Figure 2 (sketch no. 1, line no. 1, light blue versus pink frames in Figure 3).

b.) Alternation between possible and impossible textures in terms of performability. Impossible passages are indicated with orange filters. For example, sketch 3, line no. 1, yellow versus orange frames.

c.) Alternation between simple and superimposed textures (for superimposed textures, refer indicatively to sketch no. 4, line no. 2).

d.) Silences (green frames, indicatively sketch no. 2, line no. 2).

[1] "When problems have not yet been formalized, they can still be characterized by a model of computation that includes human computation. The computational burden of a problem is split between a computer and a human: one part is solved by a computer and the other part solved by a human. This formalization is referred to as the human-assisted Turing machine."

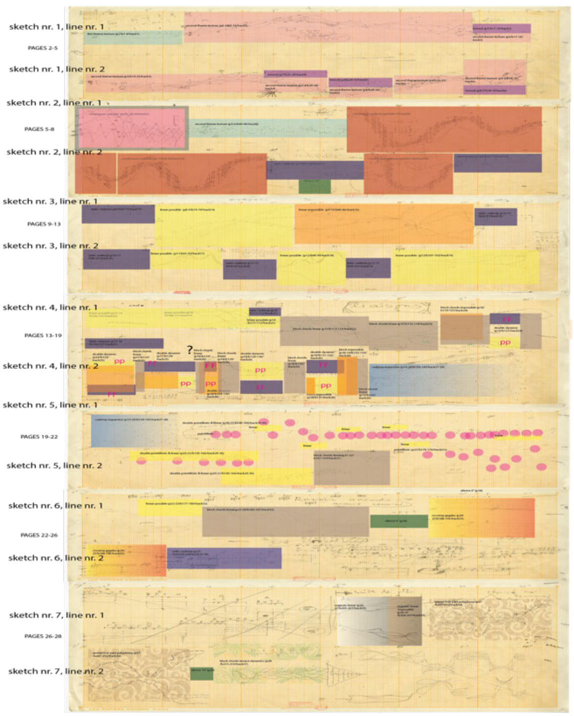


Figure 3. Complete textural and global form annotation and analysis of the sketches of *Evryali* with kind permission by the Xenakis Archive – Mâkhi Xenakis

3. ONE-TO-MANY MAPPINGS

3.1. Embodied Learning

The notion of embodied learning encapsulates a central hypothesis: that gesture/movement can be a form of processing textual complexity, including both quantitative and qualitative characteristics. This processing both reduces the dimensionality of symbolic information by folding it into higher-order co-articulation units [6, p. 2][2] and multiplies it through the production of one-to-many mappings between lower affordances, higher-order parameters or descriptors, and different hierarchical layers of embodiment.

3.2. Multimodal Data and Sketches

As is the case with symbolic scores, sketches are highly decoupled from the performers' embodied view of the work. To demonstrate this, I conducted an analysis of a performance of the "cadenza expansion" section (sketch no. 4, line no. 2 in Figure 3) using multimodal data, including Inertial Measurement Units (IMU) and MIDI, visualized, annotated and synchronized through the MuBu (Multiple Buffer) toolbox in Max/MSP. The simple triangular form of the sketch as an expansion in pitch and time space is counterbalanced by the complex hand choreography necessary to perform it. In Figure 4, this choreography is defined and visually communicated through markers indicating the position in time of hand displacements as measured by the gyroscopic information of the multimodal data. Video 1 shows an audiovisual recording of the passage synchronized to the multimodal data and subsequently an interactive demo clarifying how markers indicate displacements. The hand displacements define PADR envelopes for the pianist's gesture, P standing for zones of gesture preparation, A for zones of attacks without hand displacements, D for zones of displacements and R for the release gesture of the pianist.

The rate or density of displacements is here considered as a robust measure of complexity, which can function in relation to the ongoing research for developing complexity measures or indexes in music notation [7]. The relation between difficulty and notational complexity remains convoluted and requires further investigation.

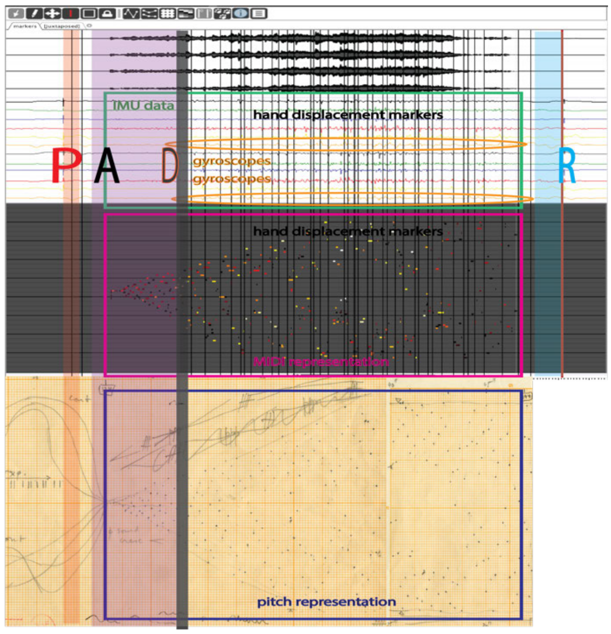


Figure 4. Comparison of a Xenakis sketch (bottom panel) with multimodal performance data including MIDI (middle panel) and IMU acceleration data (top panel). The gyroscopic information defines patterns of hand displacements that are used to define envelopes of gesture preparation (P), attack (A), displacement (D) and release (R). Sketch of *Evryali*, p.19, reproduced with kind permission by the Xenakis Archive – Mâkhi Xenakis

A good measure of the merits of Figure 4 in terms of direct perception is the citation of an annotated score of the same, as in Figure 5. In this traditional form of annotation, each displacement unit corresponds to a hand-drawn

[2] "Coarticulation means the subsumption of otherwise distinct actions and sounds into more superordinate actions and sounds, entailing a contextual smearing of otherwise distinct actions and sounds, e.g. rapid playing of scales and arpeggios on the piano will necessitate finger movements included in superordinate action trajectories of the wrists, elbows, shoulders, and even whole torso, as well as entail a contextual smearing of the singular tones into superordinate contours of the scales or arpeggios. […] One essential element of coarticulation is that it concerns both the production and the perception of sound, hence that it clearly unites sound and action into units, into what we prefer to call sound-action chunks in music".

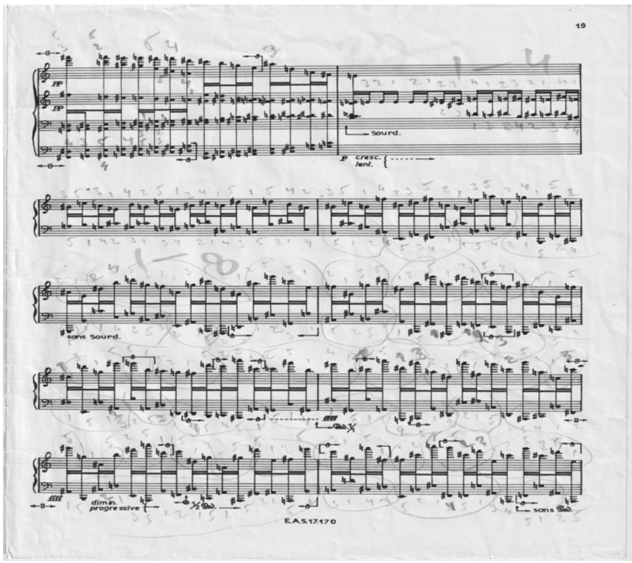


Figure 5. Published score annotation by the author for *Evryali*, p. 19, reproduced with kind permission of *Editions Salabert*

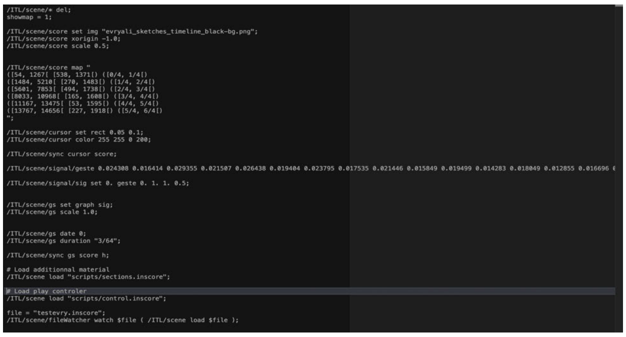
circle, indicating a hand-grasp of pitch information. While this sort of cumulative implicit knowledge remains indispensable for the performer's learning process, the neat combination of local and global aspects in Figure 4 makes explicit to the non-performer information that is absent from either the sketch or the score alone.

3.3. Augmented Interactive Tablatures

The stark contrast between local embodied detail and global formal properties in Xenakis demands an augmented interactive representation functioning as an interface for learning and performance, beyond both the massive details of the symbolic notation and the perception of form through sketches.

I employed the INScore to create an augmented interactive tablature based on a rendition of the complete *Evryali* sketches as a single timeline. Check Figure 6 for a sample of INScore's scripting language, an extended textual version of Open Sound Control messages, for the following string: /ITL/scene/score set img "evryali_sketches_timeline_black-bg.png;". This string places the representation of Figure 7 in the INScore scene.

As a first step, the INScore formalism was used to create mappings between graphic space expressed in pixels and musical time expressed in traditional time signatures ("([54, 1267[[538, 1371[) ([0/4, 1/4[)" etc). Based on this mapping, which can be variable, I synchronized elements such as cursors (/ITL/scene/sync cursor score;) and graphic signals rendering already recorded gestures (/ITL/scene/sync cursor score;).



```
/ITL/scene/* del;
showmap = 1;

/ITL/scene/score set img "evryali_sketches_timeline_black-bg.png";
/ITL/scene/score xorigin -1.0;
/ITL/scene/score scale 0.5;


/ITL/scene/score map "
([54, 1267[ [538, 1371[) ([0/4, 1/4[)
([1484, 5210[ [270, 1483[) ([1/4, 2/4[)
([5601, 7853[ [494, 1738[) ([2/4, 3/4[)
([8033, 10968[ [165, 1608[) ([3/4, 4/4[)
([11167, 13475[ [53, 1595[) ([4/4, 5/4[)
([13767, 14656[ [227, 1918[) ([5/4, 6/4[)
";

/ITL/scene/cursor set rect 0.05 0.1;
/ITL/scene/cursor color 255 255 0 200;

/ITL/scene/sync cursor score;

/ITL/scene/signal/geste 0.024308 0.016414 0.029355 0.021507 0.026438 0.019404 0.023795 0.017535 0.021446 0.015849 0.019499 0.014283 0.018049 0.012855 0.016696 0

/ITL/scene/signal/sig set 0. geste 0. 1. 1. 0.5;


/ITL/scene/gs set graph sig;
/ITL/scene/gs scale 1.0;


/ITL/scene/gs date 0;
/ITL/scene/gs duration "3/64";

/ITL/scene/sync gs score h;

# Load additionnal material
/ITL/scene load "scripts/sections.inscore";

# Load play controler
/ITL/scene load "scripts/control.inscore";

file = "testevry.inscore";
/ITL/scene/fileWatcher watch $file ( /ITL/scene load $file );
```

Figure 6. INScore script for creating an augmented and interactive representation based on Xenakis's sketches

As a second step, I experimented with different mappings and views of the score, often even with multiple renderings of the sketch simultaneously, including linear and non-linear readings. Non-linear readings allow for a navigation of the score that reflects analytical insights or embodied learning. One could, for example, group similar textures or similar choreographies, by creating respective mappings.

Figure 7 shows a rendering of the textures already presented in Figures 1 and 2, which includes a linear (upper line) and a non-linear (bottom line) reading of the sketch. Video 2 presents a screen recording without sound of a) a non-linear close-up view of the upper sketch in Figure 7 and b) the simultaneous sending of clock messages from Max/MSP to an overview of these two representations, visualized via a respective cursor and signal.

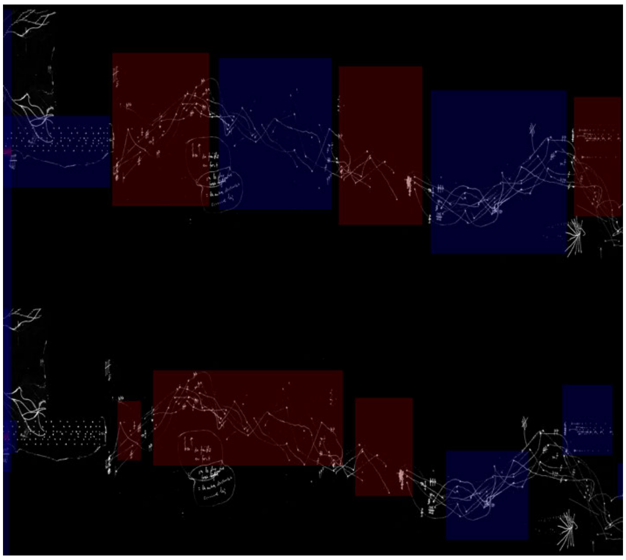


Figure 7. Graphic result of two different mappings of the *Evryali* sketches that correspond to bars 1–18. The upper mapping is linear, thus the alternation of blue and red regions defines the mapping. The lower mapping is non-linear, connecting any regions in the sketch, thus the succession of monochromatic regions.

4. FROM SCORE REPRODUCTION TO STRUCTURED COMPROVISATION

Having already learned the piece in meticulous detail, the augmented interactive tablature with variable representations and mappings operated as a platform for initial experimentation with navigating similar textures, but also harmonic and melodic elements of the original. At this stage, I simply followed the mobile elements of the tablature along predefined paths.

In what follows, different ways of using the sketch-based augmented multimodal tablature for structured comprovisation are explored. First, I explore ways of training the tablature to follow the performers' gestures and applications in solo comprovisation. Then, I explore a duo comprovisation involving the AI agent SOMAX2 controlled by another human performer and the trained interactive tablature.

4.1. Training the Tablature: Motion Follower

By virtue of a syntax of movement and machine learning techniques, the multilayered tablature can be trained to follow the performer in variations of the initial performance. This system is based on a probabilistic motion-following methodology employing Hidden Markov Models [8] and on the PADR envelopes demonstrated in 3.2. The crucial element, that allows for the motion-following to be reflected in the notation and thus become score-following, is that both the gesture and the notation share the same basic segmentation.

The process involves a *recording phase* and a *following phase*. In the recording phase, the user follows any mobile element of the INScore, which is set to move at a desired speed as in Video 2, like a classic metronome would do. The musical sketch has already been graphically segmented and assigned a duration according to the INScore space-time formalism (explicit mapping). In this phase, the motion follower "learns", so to speak, the mapping from the performer's gesture captured by R-IoT IMUs (implicit mapping), while s/he follows the mapping of the INScore (explicit mapping). In the next phase defined as "following", the performer can pursue highly varied performances, ranging from heterophonic re-interpretations of the original to the introduction of a completely novel material that shares the same gestural segmentation. This time, it is not the performer that follows the system, but rather the system that follows the performer, given that the segmentation is correct and common in all these varied performances. Thus, the performer may control the mobile elements of the INScore tablature. The feedback of the follower has been extended to score compound representations. The gesture-following has been turned into score-following.

In Figure 8 the grey signal represents the implicit mapping gesture that the augmented multimodal tablature "learns" along the explicit mapping of Figure 7 in the recording phase. The green signal represents the new incoming signal that controls the tablature in the following phase, the signal that the tablature follows during variations of the initial performance.

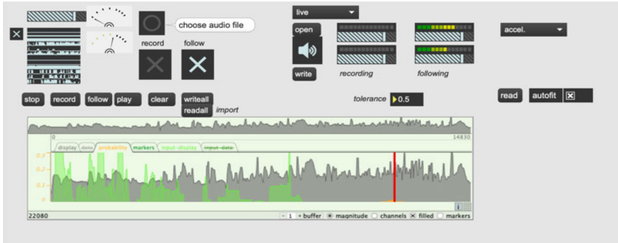


Figure 8. The green signal represents a new gesture that is probabilistically compared to the already recorded grey signal, allowing for the following of the performance by the system according to a threshold of tolerance

4.2. Training an AI Agent for Comprovisation: SOMAX2

In the last phase, I combined the motion following sketch-based tablature with a corpus of *Evryali* audio recordings used as training material for SOMAX2, which was trained and controlled live by Mikhail Malt. According to Malt, SOMAX2 is a multi-agent interactive system performing live machine comprovisation with musicians, based on machine-listening, machine-learning, and generative units. The actual version [9] is a recent development and algorithms' improvement from the former SOMAX version. Agents provide stylistically coherent improvisations based on learned musical knowledge while continuously listening to and adapting to input from musicians or other agents in real time. The system is trained on any musical materials chosen by the user, effectively constructing a generative model (called a corpus), from which it draws its musical knowledge and improvisation skills. Corpora, inputs and outputs can be MIDI as well as audio, and inputs can be live or streamed from MIDI or audio files. SOMAX2 is one of the improvisation systems descending from the Omax software [10], presented here in a totally new implementation. As such it shares with its siblings, the general loop [listen/learn/model/generate], using some form of statistical modeling that ends up in creating a highly organized memory structure from which it can navigate into new musical organizations, while keeping style coherence, rather than generating unheard sounds as other ML systems do.

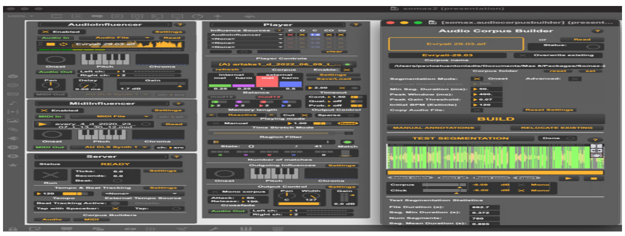


Figure 9. Central window of SOMAX2 featuring the segmentation of an audio file containing a recording of *Evryali* (right side in green)

4.3. Improvising with the Tablature and SOMAX2

The open-ended nature of the resulting system allowed us to experiment with different kinds of interaction and materials in a setting of structured comprovisation based on Xenakis's *Evryali* sketches. Capitalizing on the *chroma* affinities between Xenakis's complete piano works, as well as music by Ravel and Janáček, we were able to construct composite corpora including harmonically similar language. In Videos 3,4,5 we present three different instances.

In the first instance (Video 3), the comprovisation between the human pianist (the author) and the AI agent controlled by Mikhail Malt is based exclusively on material by *Evryali,* featuring several degrees of distancing (same, similar and alien material) of the human pianist from the original material used to train SOMAX2. The complete performance is documented here.

In the second instance (Video 4), the pianist performer is gesturally controlling a recording from Xenakis's *Mists* (1980), first with air gestures and then with new material on the piano. The new material is shaped by the performer according to an *Evryali* tablature (Figure 10, pink frame). The supervp.scrub~ object (advanced phase vocoder position controlled player module) allows resynthesized audio output from the follower to be sent to SOMAX2, still controlled by Mikhail Malt, improvising on the time-stretching, pitch transposition, spectral envelope transformations of the original recording at times when the systems fail to follow.

In the third instance (Video 5), the piano performer is similarly performing air gestures and a heterophonic duet based on audio material from "Oiseaux Tristes" by Maurice Ravel, his movements controlling the *Evryali* tablature projected in the right-hand corner of the video (Figure 11). The SOMAX2 controlled by Mikhail Malt responds as above (instance 2).

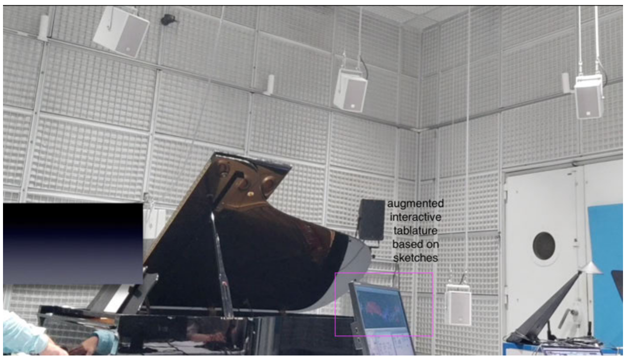


Figure 10. Snapshot from Video 4, annotated. Augmented interactive tablature based on *Evryali* sketches is used as a graphic score for a co-adaptive improvisation of a human performer, Pavlos Antoniadis, with SOMAX2 controlled by Mikhail Malt, including air gestures and piano based on another piece by Xenakis, *Mists*

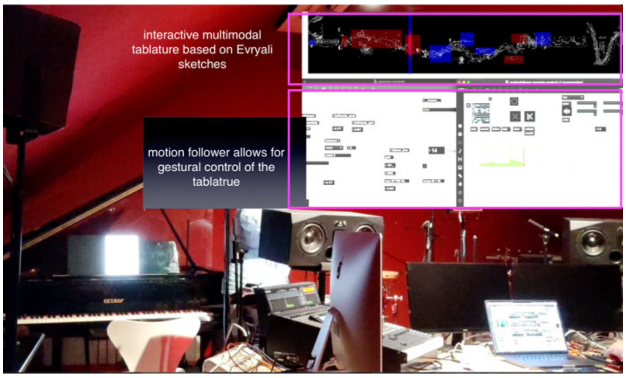


Figure 11. Snapshot from Video 5, annotated. Augmented interactive tablature based on *Evryali* sketches is used as a graphic score for a co-adaptive improvisation of a human performer, Pavlos Antoniadis, with SOMAX2, controlled by Mikhail Malt, including air gestures and piano based on Maurice Ravel's "Oiseaux Tristes"

5. SKETCHES AND EPISTEMIC BIASES IN ARTIFICIAL INTELLIGENCE

The reproduction of implicit and explicit human biases in algorithms has been a hot topic in AI research. Ethical questions about open AI applications have recently focused on discriminatory socio-political biases such as gender, race [11] and the Global North [12], whereas fundamental questions about implicit or explicit epistemic biases in relation to knowledge representation in AI have been less prominent [13]. While these two categories are fuzzy and far from incompatible [14], epistemic biases tend to encompass an endless range of *a priori* conceptions and models of human understanding, learning, knowledge, and judgment, which are uncritically implemented in AI applications [15]. As a result, systems architectures, as well as human-in-the-loop components, are constrained by traditional models of knowledge and fail to embrace the latest paradigm shifts in their target domains.

Both probabilistic architectures employed in the examples above, the gesture follower and the SOMAX2, exhibit strong implicit epistemic biases as far as knowledge representation in music is concerned. Their mutual reliance on the probabilistic navigation of audio files, either in relation to an incoming gesture in the case of gesture-follower, or in relation to pitch and *chroma* characteristics in SOMAX2, resonates with what has been called "the semantic blind spot of current inferential accounts of AI" [16]: Parmenidean probabilistic syntactic interchangeability of memoryless states produces no semantic relations or illusion of real effects of causality, unless *real* Heraclitean change is effectuated. And real change is effectuated by virtue of the unknown, rather than by virtue of known navigable corpora.

In retrospect, one could further claim that the implicit symbolic biases of the classic reproduction model, or even of its updates in forms such as the *high-modernist model of performance practice* [17], including: the fixity of musical scores in space and time; their parametric stratification; the symbol grounding problem [18] of musical parameters and *the inside – outside of time* problem that Xenakis and others have uttered, are transposed in straight-jacket fashion to a poor rendition of a complex dynamic system of improvisation [19], with biases such as: the limits of symbolic representation in relation to embodied experience "escaping computation"[3], the privileging of abstract sound relationships codified in scores over multimodal interactions embedded in social contexts, and the dimensionality reduction of complex and dynamic stances to simple parameters.

And yet: the concept of an emergent multiplicity of mappings between decoupled representation domains – sketches, scores, and multimodal data – is a promising one, in that it allows for unpredictable and personalized meaning-producing *inside time* interactions, even when the respective domains may seem to be highly sophisticated and responsive reshufflings of fixed *outside of time* timelines. The prospect of cracking the one-to-one mappings of sketches open, via their implicit one-to-many mappings and through their controversial existential repurposing into graphic scores affording similar and alien materials, is tempting, if only provisional, as any guerrilla tactic of malleable sketching should be.

6. CONCLUSIONS

Following a bottom-up trajectory, I have attempted to present a methodology that leads from deciphering and learning processes based on a meticulous analysis of sketches aiming at high modernist performance practice, to the open-ended use of sketches as graphic scores in Human-Machine comprovisation settings, including machine learning and Artificial Intelligence techniques. Such methodology reveals the inherent transitoriness of sketches as media, but also the importance of fluid forms of knowledge representation and of Human-Assisted Turing Machines amidst current trends and media-hypes in AI, which accentuate the blackboxness of Deep Learning based on Big Data and the "AI Effect" that renders "lighter" forms of AI obsolete, absorbing them into the broader category of computation. Featuring Iannis Xenakis's music for this purpose is far from random, given the dialectic clash between strict formalism and sensational surface, body and mind, scientific and artistic research epistemologies of his output.

Acknowledgments

I would like to warmly thank: The Association Les Amis de Xenakis, Mâkhi Xenakis, and Makis Solomos for kindly providing access to Xenakis's sketches for *Evryali*; Dominique Fober for his advice on INScore programming; Frédéric Bevilacqua, Riccardo Borghesi and the *interaction-son-musique-movement* team at IRCAM for the long-term collaboration on gesture capture; Frédéric Bevilacqua and Elaine Chew for conducting and facilitating multimodal recordings at Ircam in the context of the COSMOS project (September 2020); Mikhail Malt, Marc Chemilier and Gérard Assayag for their kind invitation to the "Journée d'études en hommage à Iannis Xenakis", Jeudi 12 mai 2022 at the Auditorium de l'ISC-PIF, Paris and to the *Improtech* Workshop – Festival in Uzeste, France 11–13 August 2023. This project was partly funded with the support of ERC Advanced Grant REACH (Raising Co-Creativity in Cyber-Human Musicianship).

7. REFERENCES

[1] I. Xenakis, *Formalized Music*, Hillsdale: Pendragon Press, 1992.

[2] D. Exarchos, "Outside of Time," in *A Xenakis Dictionary*, ed. D. Exarchos, Association Les Amis de Xenakis, 2023. [Online]. Available: https://www.iannis-xenakis.org/en/outside-of-time

[3] N. Cook, *Beyond the Score: Music as performance*, Oxford University Press, 2013.

[4] D. Shahaf and E. Amir, "Towards a Theory of AI Completeness," in *Proc. 8th international Symposium on Logical Formalizations of Commonsense Reasoning*, Stanford, 2007. [Online]. Available: www.ucl.ac.uk/commonsense07

[5] J. Harley, "Arborescence," in *A Xenakis Dictionary* ed. D. Exarhos, Association Les Amis de Xenakis, 2023. [Online]. Available: https://www.iannis-xenakis.org/en/arborescence

[6] R. I. Godøy, "Understanding Coarticulation in Musical Experience," in M. Aramaki, O. Derrien, R. Kronland-Martinet, S. Ystad (Eds.), *Sound, Music, and Motion 10th International Symposium*, CMMR 2013 Marseille, France, October 15–18, 2013

[7] S.C. Chiu and M.S. Chen, "A study on difficulty level recognition of piano sheet music," in *Proceedings of the 2012 IEEE International Symposium on Multimedia*, ISM '12, Washington DC, 2012, pp. 17–23

[8] F. Bevilacqua, N. Schnell, N. Rasamimanana, B. Zamborlin, F. Guedy, "Online Gesture Analysis and Control of Audio Processing," in J. Solis and K. Ng (eds.), *Musical Robots and Interactive Multimodal Systems*. Berlin: Springer-Verlag, 2011, 127–142.

[9] J. Borg, *Dynamic Classification Models for Human-Machine Improvisation and Composition*, Master's report, Ircam and Aalborg University Copenhagen, 2020.

[10] G. Assayag et al., "Omax brothers: a dynamic topology of agents for improvisation learning," in *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*. Santa Barbara, CA, 2006, pp. 125–132.

[3] https://provocations.online/whatescapescomputation

[11] L. Marinucci, C. Mazzuca and A. Gangemi, “Exposing implicit biases and stereotypes in human and artificial intelligence: state of the art and challenges with a focus on gender,” *AI & Soc*, 2022. https://doi.org/10.1007/s00146-022-01474-3

[12] M. Donnarumma, “AI Art Is Soft Propaganda for the Global North”, 2022. [Online]. Available: https://hyperallergic.com/772848/ai-art-is-soft-propaganda-for-the-global-north/?fbclid=IwAR0S1R7F7w0a3yqTzWP4zCLR8L08T6BWjyQf9I2K6qOpaPpG5ELB5mvyOc8

[13] M. Shanahan, “Talking About Large Language Models,” arXiv:2212.03551v2 [cs.CL], 2022. https://doi.org/10.48550/arXiv.2212.03551

[14] E. Beeghly and A. Madva (eds.), *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind,* New York: Routledge, 2020. https://doi.org/10.4324/9781315107615

[15] R. Negarestani, *Intelligence and Spirit*, Falmouth: Urbanomic, 2018.

[16] AA. Cavia, *Logiciel. Six Seminars on Computational Reason*, Berlin: The New Centre for Research and Practice, 2022.

[17] F. Cox, “Notes Toward a Performance Practice for Complex Music,” in Mahnkopf, C.-S., Cox, F. and Schurig, W. (eds.): *Polyphony and Complexity, New Music and Aesthetics in the 21st Century*, Hofheim: Wolke, 2002, pp. 70–132.

[18] L. Barsalou, “Perceptual Symbol Systems,” *Behavioral and Brain Sciences* 22 (1999), pp. 577–609. https://doi.org/10.1017/S0140525X99002149

[19] D. Borgo, *Sync or Swarm. Improvising Music in a Complex Age*, Bloomsbury, 2022.

FROM SONIC GESTURE TO COMPOSITION: MAPPING THE PATH FOR THE PHYSICAL COMPUTING INSTRUMENT

João Coimbra
Aveiro University/INET-md
Aveiro, Portugal
jpcoimbra@ua.pt

Henrique Portovedo
Aveiro University/INET-md
Aveiro, Portugal
henriqueportovedo@ua.pt

Sara Carvalho
Aveiro University/INET-md
Aveiro, Portugal
scarvalho@ua.pt

ABSTRACT

As a composer and multi-instrumentalist, I use the studio as a tool for real-time performance and composition. In recent years, this process has suffered constraints due to my need to incorporate more studio elements to produce and manipulate sound. My research project seeks to address these challenges by developing and mapping a digital musical instrument – the Physical Computing Instrument (PCI) – that uses gestures to produce sound.

In this paper, we delve into recent developments in our research, exploring three gestural Digital Musical Instruments (DMIs) as exploratory tools for refining the PCI's design and functionality. We focus on the musical challenges arising from it, encompassing idiomatic writing, notation, and performance, and how these aspects contribute to the PCI's development, clarifying its scope and identity, while establishing a systematized approach for its practice.

Additionally, the paper introduces two original compositions, accompanied by notational sketches and mapping strategies tailored for these gestural controllers. These elements provide valuable insights into the prototyping of the PCI, showcasing its potential applications.

1. INTRODUCTION

My artistic practice hinges on performance as a real-time composition process, where movement and gesture convey instrumental meaning [1]. The music I craft is intrinsically tied to instrumental performance. It is the performance of sound that leads me to composition. This process makes use of traditional, analog, and digital instruments, as well as microphones, sensors, DACs, MIDI interfaces, and notational software, promoting the interaction of elements to favor sound-based composition. This dynamic process has suffered constraints due to my need to augment the studio with more elements to produce and manipulate sound. By constraints, I allude to the growing challenge of maintaining the immediacy of what Van Nort aptly termed "action-sound coupling" [2]. To surmount these constraints, I propose the introduction of the Physical Computing Instrument, an addition to my studio and my live performances. The PCI is envisioned as an instrument that generates sounds and seamlessly connects and interacts in real-time with existing studio elements. The PCI will be a sound-producing device that can be controlled by a variety of physical gestures and is reactive to user actions [3].

One of the main challenges regarding the creation of a digital instrument with gestural control is the establishment of the artificial relationships that govern sound production and its control. As d'Escriván points out, this is a novel frontier with contours yet to be fully defined [4]. Jordà states that a new instrument must be capable of shaping thinking, relations, interactions, and temporal and textural organization, ultimately leading to the emergence of new kinds of music [5]. If challenges are extended to immersive visual and sonic environments, there is also a need to address the rich and autonomous ecosystems within these environments, as Hamilton has pointed out. These ecosystems have their own rules and realities that can impact user interactions [6].

The challenges extend to notation, as there is no standardized system for these new instruments. Thus, in the DMI domain, notation could be embedded in the instrument. As Thor Magnusson points out, pieces could end up having a notation that is not specific to the instrument but refers to the type of mapping it needs [7]. Notation beyond its intrinsic value – a code to compose and to perform – could serve multiple purposes, including acting as a design constraint for new digital instruments, as it can lead to a systematization and instrumental practice. Moreover, a written repertoire both for solo and ensemble contexts may improve the DMI's longevity while attracting a broader community [8]. In that context, the PCI seeks to achieve a dynamic relationship between its physical affordances and idiomatic notation to scale up its full potential as a tool for new repertoire. It is also intended that this new repertoire could be written without the need for special vector graphics software, instead using readily available symbols in standard notation software like *Finale* or *Sibelius*.

Moreover, the PCI is designed to be accessible to non-specialized practitioners who may not be familiar with technology but are eager to expand their artistic horizons. It must be user-friendly and feature notation that is easily comprehensible, leveraging symbols and layouts familiar to practitioners. It features a gentle learning curve, portability, and the capability to augment existing instruments. This aspiration aligns with the design principles of acoustic instruments, considering factors like ergonomics, range, expressiveness, instrumental gestures, durability, and mapping choices [9].

To illuminate these concepts, I employ three gestural controllers as probes – the Leap Motion V.1, an optical hand-tracking module that captures the intricate movements of hands and fingers; Wii Remote & Nunchuck Motion Plus, the primary controllers for the Nintendo Wii, and Myo Armband, a gestural armband that reads biosignals (Electromyographic signals – EMG) and remotely connects with other devices. In the following sections, I will present two compositions: "Tiny Oracle for Wii Remote and Leap Motion", and "Tiny Oracle for Solo Flute and Myo Armband". These compositions exemplify how these gestural controllers serve as valuable research tools, paving the way for the implementation of the PCI and the development of idiomatic notation.

2. CONTEXT

Central to my artistic research is the use of timbre as a fundamental compositional tool. Given my focus on sound processing, synthesis, and the creation of non-acoustic sounds, the instrument I aim to construct falls within the domain of Digital Musical Instruments. The *Physical Computing Instrument* consists of two primary elements: a gestural controller and a sound generation unit. The gestural controller, or gestural interface, will manage and transmit data encoding both performer actions and composer intentions. It will incorporate an array of sensors, including bio-signal sensors (such as electro-myographic), motion tracking sensors (infra-red, Nine degrees of freedom – 9DoF), haptic sensors, and wireless sensors for seamless data exchange between the two units.

All this information will be mapped into a sound generation unit, the second element of the instrument, and decoded in a MaxMSP patch. Beyond sound generation and the mapping of bodily parameters, this patch will facilitate communication with studio elements (such as DAW, analog synthesis, and signal processing), expanding the instrument's possibilities. Given its highly technological nature, collaboration with the Department of Materials and Ceramics at the University of Aveiro (CICECO) has been sought for the implementation and construction of the instrument as well as the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC).

The current stage of the *PCI* implementation comprises the following items:

- Gestural Implementation through existing DMIs: *Leap Motion*, *Wii Remote*, *Myo Armband.*
- Pieces and compositional sketches, combining the three DMIs and traditional instruments.
- Study of mapping possibilities and signal acquisition.
- Construction of a prototype.
- Notational studies.

An initial concept sketch of the instrument is provided in Figure 1.

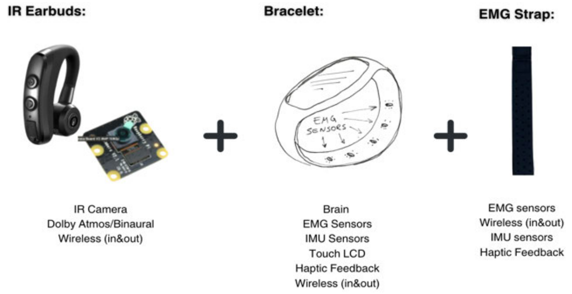


Figure 1. Physical Computing Instrument (Sketch)

The PCI comprises three essential components: IR Earbuds, a Bracelet (the brain), and an EMG strap. The infrared (IR) earbuds offer versatile practical applications, serving as a component in tracking hands and finger movements. As an example, it can be employed to train models that will assist a performer in changing parameters like sounds and patches remotely, providing additional layers of musical expression. We are testing this possibility using Machine Learning (ML) software like *Wekinator* or *ml.lib* a cross-platform for MaxMSP and Pure Data [10, 11]. Additionally, ML models could be used to facilitate remote page turning in digital readers, freeing performers from the constraints of physical scores or using trigger pedals for that function.

Another feature of the earbuds is the spatial capabilities, decoding 360º formats, propelling the PCI into the realm of real-time performance and composition within a 3D environment. This feature transforms the instrument into a versatile tool for both studio work and live performances, given the provision of a spatial system. A simple arm gesture (using the inertial movement unit – IMU at the bracelet) becomes a conduit for directing sound to any desired point in a room, enhancing the spatial dynamics of the musical experience. This real-time feedback mechanism not only enhances the performer's control over the spatial aspects of the performance but also ensures an immersive and precisely orchestrated auditory experience.

The bracelet itself is designed to be a comprehensive control center for the PCI. Linked wirelessly to the earbuds and the EMG Strap, allows the mapping of the performer's arms, hands, and fingers. It also integrates an LCD touchscreen that serves as an intuitive interface. This touchscreen enhances the accessibility of the instrument's features but also provides real-time visual monitoring of its operation. The inclusion of this visual element minimizes the reliance on external devices, effectively mitigating the need for a laptop during performances.

3. DMIs AS PROBES

None of the three interfaces that I chose as a research path for the PCI's implementation were conceived as musical instruments. However, their latent potential has been harnessed and expanded upon by the NIME (New Interfaces

For Musical Expression) community. Its mapping contained gaps that led to the implementation of alternatives for its use in music, such as the *Myo mapper* or the Myo-to-Osc developed for the 'RAW' instrument, a platform for collective performance [12, 13]. Similarly, the *Leap mapper* app is designed for the *Leap Motion* [14].

The three gestural controllers operate on distinct technologies, providing diverse mapping and feature possibilities. These and other software alternatives facilitate communication into various protocols such as MIDI or OSC for practical application with the interfaces at my studio [15]. Next, I'll show two compositions that exemplify how these gestural controllers are being used as valuable research tools, paving the way for the implementation of the PCI and the development of idiomatic notation.

3.1. Tiny Oracle – For Wii Remote and Leap Motion

Tiny Oracle for Wii Remote and Leap Motion (2023)[1] is the first piece that I wrote within the scope of the research. The sound materials result from 20 prepared piano samples processed with granular synthesis, pitch, and phase shifting techniques.

Figure 2. Recording Setup at DECA's Studio (INET/md)

Although I didn't write it using the two DMIs – I used a MIDI keyboard – the musical gestures were performed having in mind the affordances of the *Wii remote* and the *Leap Motion.* After writing the piece, I started the implementation of the piano samples in the *Wii Remote* and the *Leap Motion* using *OSCulator*, *Leap Mapper*, and an *EX24* soft sampler.

The implementation consists of mapping the sounds to be playable by the gestural controllers and inferring if their affordances are satisfactory from the performer's point of view [16]. The refinement of performance expression, technique, and ergonomics becomes a focal point during this stage.

Questions surface organically, steering the exploration towards the heart of artistic feasibility, such as "Is it possible for these frugal interfaces to interpret and perform the piece effectively?" and "Can the sounds produced by these controllers be accurately represented in a musical score?"

/hand1/pinch	MIDI Note	F#1
/hand1/position	–	–
0	MIDI Note	C0
1	MIDI Note	G1
2	MIDI Note	C3
/wii/1/accel/pry	–	–
0: pitch	MIDI CC	1
1: roll	MIDI CC	1
2: yaw	MIDI CC	2
3: accel	MIDI Note	G#1
/wii/1/button/1	MIDI Note	F#1
/wii/1/button/2	MIDI Note	C#1
/wii/1/button/A	MIDI Note	A1
/wii/1/button/B	MIDI Note	D1
/wii/1/button/Down	MIDI Note	F2
/wii/1/button/Home	MIDI Note	C1
/wii/1/button/Left	MIDI Note	F1
/wii/1/button/Minus	MIDI Note	F3
/wii/1/button/Plus	MIDI Note	A0
/wii/1/button/Right	MIDI Note	F0
/wii/1/button/Up	MIDI Note	D0
/wii/1/nunchuk/accel/pry	–	–
0: pitch	–	–
1: roll	–	–
2: yaw	–	–
3: accel	MIDI Note	A3
/wii/1/nunchuk/button/C	MIDI Note	C5
/wii/1/nunchuk/button/Z	MIDI Note	C#0
/wii/1/nunchuk/joy	–	–
0: x	MIDI Note	C6

Figure 3. Sampler mapping using OSCulator

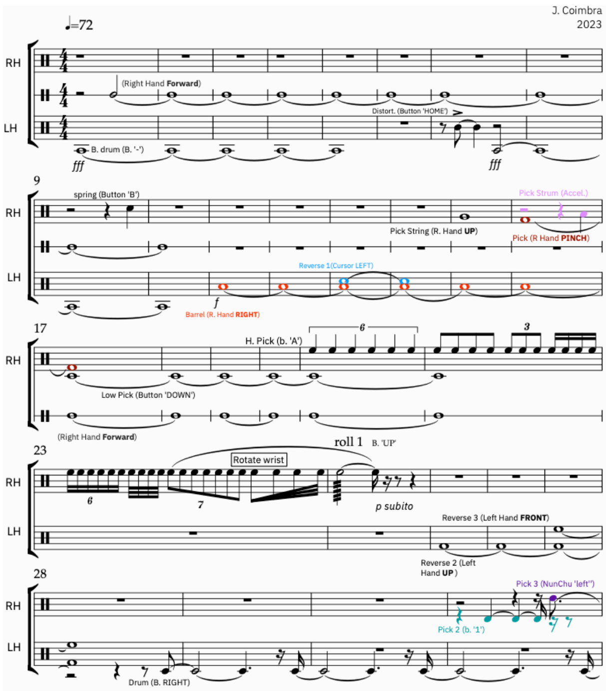


Figure 4. Tiny Oracle for Solo DMI (excerpt)

For notating *Tiny Oracle for Wii Remote and Leap Motion V.1,* I opted for a grand staff with a percussion-based stave in between. I spread the high register samples at the upper stave (right hand) and the low register samples at the below stave (left hand). The performer uses both hands to control the gestural controllers.

[1] A sound file of the piece it's available under: https://on.soundcloud.com/yr1mt

The percussion-based stave was intended to notate the triggering of a specific sound, using neck and wrist rotation (for a continuous manipulation of timbre), but that proved impossible to map with these controllers. The idea was to free the hands to the sounds that demand a rhythmically precise performance and use the neck to trigger sounds that change over time and that aesthetically work as a sound metaphor for the human body, specifically the cervical vertebrae [17].

As an alternative, I have mapped that specific sound to be triggered by *Leap Motion* with the instruction "Right Hand Forward" (see Bar 1 in Figure 4.).

Figure 4 visually articulates the distribution of samples across the grand staff and the percussion-based stave, illustrating the thoughtful design considerations made to navigate the unique challenges posed by the gestural controllers. Additionally, colored note-heads have been used to check if it helps in memorization.

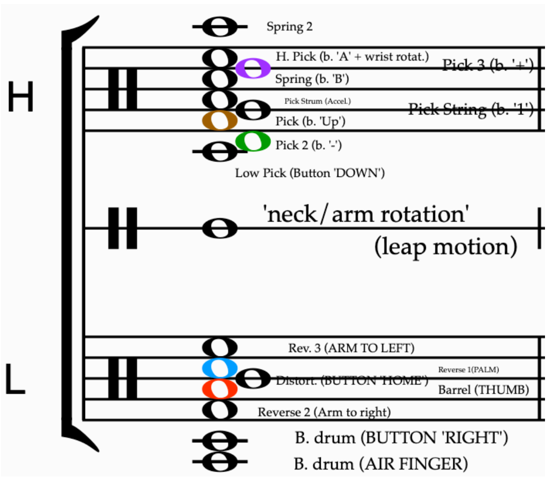


Figure 5. Tablature

I played the piece live recently. The choice of a traditional grand staff, despite not correlating with precise pitch, helped as a practical approach to score memorization [18]. Even in the absence of a direct correspondence between symbol and pitch, the arrangement of samplers by the register in a traditional score provided a mnemonic aid. This structuring also enabled the precise notation of rhythm, aligning with the demands of my sound-based composition [19]. I played the piece using the following setup: *Wii Remote*, *Nunchuck*, *Leap Motion*, *Osculator*, *Leap Mapper*, an *EX24* Sampler, a laptop, and a *DAC* interface[2]. I then did an improvisation piece using the same setup plus a *soft synth*, and an amplified cymbal. It was surprising to witness how these frugal interfaces effortlessly conveyed significant instrumental meaning through the expressive movements and gestures they enabled.

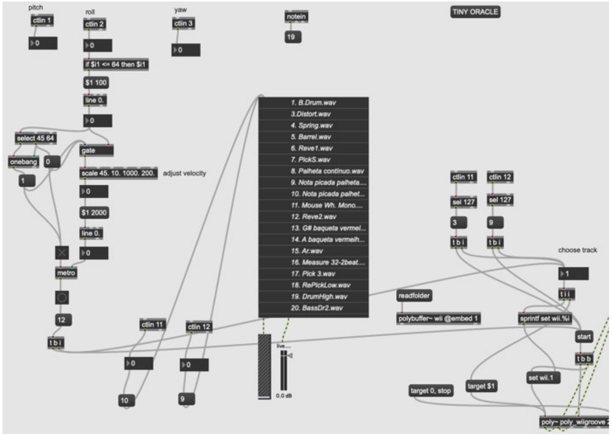


Figure 6. Tiny Oracle MaxMSP implementation

Figure 7. Performance at jornadas CICECO'23

We're already implementing the piece in MaxMSP since it will be the platform chosen for the prototype due to its mapping flexibility.

3.2. Tiny Oracle – For Flute and Myo Armband

Tiny Oracle for Flute and Myo Armband[3] is an adaptation of *Tiny Oracle for Solo Flute* a commission for the "Frederico Freitas Award" of Aveiro University. Again, I didn't write the electronic making use of the gestural controllers, but the musical gestures were created having in mind the affordances of the *Myo Armband* [20].

This piece introduces the utilization of electro-myographic sensors (EMG) and the Inertial Measurement Unit (IMU) of the *Myo* to dynamically trigger sounds. The activation of the samplers is achieved through an 'air fingering' technique, engaging the *EMG* sensors, and through the performer's movements that activate the unit's accelerometer. The captured data is then transmitted by *Myo* via *Bluetooth Le* to a computer, where the sounds are stored.

[2] A video of a rehearsal and an interview explaining the implementation of the piece is available under: https://youtu.be/Qn6SXdVpcqU.

[3] A sound file of the piece is available under: https://on.soundcloud.com/iDBzY.

The *Myo* device is worn on the performer's forearm, specifically on the right forearm for ergonomic considerations.

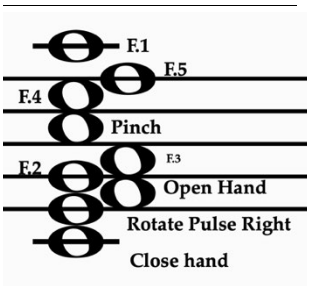


Figure 8. Tablature of Myo gestures

The score is structured with two staves, with the upper stave dedicated to *Myo* and the lower stave notating the Flute. In the context of *Myo* notation, 'X' note heads represent sound triggering, continuous control of timbral manipulation, and three-dimensional sound spatialization using *ZHDK ambisonics* externals in MaxMSP.[4] 'Diamond' note heads represent sound triggering in the vertical plane (up and down) and continuous control of timbral manipulation with spatial correlation granted by the *ambisonics* system. Standard note-heads signify sample triggering at specific points in the timeline.

The triggering of the samplers represented by standard note-heads is accomplished with gestures such as lifting fingers from the flute keys to a sufficient distance to execute the ‘air fingering’, opening, closing, and pulse rotation clockwise using the right hand.

Samplers represented by ‘X’ and ‘diamond’ note heads are triggered by the player’s movement and make use of the accelerometer and the IMU unit of the gestural controller.

An excerpt of the score, presented in Figure 8, provides a visual representation of these notational instructions.

All samplers are programmed with ‘Note In’ and ‘Note Off’ commands, providing precise temporal control and flexibility. Taking bars 21, 23, or 29 as examples, one can observe varying durations of 'X' notes. At the same time, the samplers are subject to signal processing induced by the player's movements, resulting in timbral manipulations. ‘X’ and ‘diamond’ notes mapping not only trigger sounds but permits their manipulation during the notated musical gesture.

The mapping process is executed through MaxMSP, using the *Myo Mapper* to convert *Myo* data into *OSC* messages. Additionally, a foot pedal is employed to activate and deactivate the *Myo*, preventing inadvertent triggering.

The piece’s implementation is being realized in collaboration with a flutist, who contributes valuable insights to establish a cohesive connection between the instrument, the prescribed musical gestures, and the electronic implementation. Using 'extended' traditional notation enhances sight-reading, allowing the musician to concentrate on challenging aspects of the score while providing more time for focused practice. Mastery of this piece involves not only learning the flute score but also understanding the intricacies of the electronic setup, including *Myo*, *Myo Connect*, *Myo Mapper*, *Max* patch, and a 9-channel *DAC* for 2nd order *ambisonics*. This collaborative effort is an opportunity to familiarize the performer with the live setup system [21]. The increasing familiarity with the four elements – score, *myo connect*, myo mapper, and *MaxMSP* – will help perform the piece more easily and may lead to developing a richer, more personal musical language with which to interpret it [22]. Limiting the complexity of the score is also an auto-imposed constraint that intends to free the interpreter to interpret.

Figure 9. Tiny Oracle for Flute and *Myo* (Excerpt)

Finally, the piece deals with the interaction of the acoustics proprieties of the room/concert hall where it will be performed. The player moves on stage stopping at various spots (“Stations”) and listening to the changes in sound in between. The audience is invited to listen and be aware of those changes that blend the “natural” acoustic of a particular venue and the emission of the electronic musical material provided by the *ambisonics* environment.

To conclude, using DMIs as probes allows us to assess how these pieces can be played and what is possible to do using the frugal technology provided. Thus, it elucidates how to develop idiomatic notation not only within the scope of these DMIs but also for the PCI.

[4] https://www.zhdk.ch/forschung/icst/software-downloads-5379

4. NEXT STEPS

The forthcoming phase of this research encompasses the construction of a prototype in collaboration with the Materials Department at Aveiro University (CICECO). This collaborative effort aims to synthesize the theoretical groundwork into a tangible prototype, setting the stage for further exploration and refinement.

Integral to this process is creating new musical pieces, delving into the nuances of musical gestures, technological implementation, and musical notation. Concurrently, the research into self-powered interfaces remains a focal point. In partnership with CICECO, experimentation with energy harvesting techniques is slated, exploring possibilities such as extracting energy from the human body, harnessing solar power, and capitalizing on the instrument's performance-induced vibrations [23, 24].

The research trajectory will continue to involve close collaboration with performers, engaging them in the implementation of the pieces. Furthermore, a comprehensive inquiry is underway, targeting performers and composers with diverse backgrounds. The survey navigates their relationship with technological tools in both performance and composition, encompassing aspects such as preferences between fixed-media and real-time performances, possibilities for timbre augmentation on traditional instruments, sight-reading, and feedback experiences of triggering electronic sounds through metaphorical gestures. Insights into the execution of scores with real-time electronics on stage, coupled with participants' expectations for features to be integrated into the PCI, will inform the ongoing evolution of this research.

The ultimate version of the *Physical Computing Instrument* prototype will be utilized in performances, not only for premiering new compositions with diverse ensembles but also for revisiting the mapping and, in some cases, the composition of the pieces discussed in this paper. This is necessary as certain musical gestures couldn't be fully realized with the three Digital Musical Instruments presented earlier.

5. CONCLUSION

My doctoral project aims to increase the possibilities of sound-based composition, resorting to the implementation of a new digital instrument of gestural control. The *Physical Computing Instrument* will optimize my sound-based composition as a new element at my studio while maintaining the real-time manipulation of its current elements. This instrument for sound creation through synthetic processes will be explored in the composition of several works with different instrumental formations that will help shape its features and final design.

I have elaborated on the use of existing designs of gestural DMIs such as the *Leap Motion,* the *Wii Remote*, and the *Myo armband,* arguing that their different technological characteristics could contribute to understanding the possibilities of gestural implementation, notation, mapping strategies, and other features, in the PCI's construction. I presented pieces and notational sketches that exemplified how I'm using these gestural controllers as a research path.

Future developments in the project include the construction of the PCI whose sketch was presented in this paper and several pieces and performances that will serve as design constraints, improving technological implementation, and notation, refining my gestural sound-based composition in the context of contemporary orchestration. Demonstrating the PCI's potential in different configurations of instrumental ensembles and different artistic practices through the composition and performance of novel repertoire sketches and pieces could contribute to delivering use cases to a broad community, including researchers, composers from different backgrounds, trained and not-trained musicians and to those not familiar with technology but that would like to experiment an instrument that it's intended to have a minimal additional cognitive load, that is portable, simple to use, affordable, and that could expand their artistic practice.

It is anticipated that the notation will catalyze the engagement with the Physical Computing Instrument. Using an 'extended' traditional notation whenever possible aids musicians in addressing other aspects of the score and may expedite their familiarity with it, facilitating the development of a more nuanced and personal interpretation. As already referred, the performances will play a crucial role in studying the PCI's behavior in both concert and studio contexts, yielding valuable data on the pieces, the implemented technology, and the diverse forms of interaction it generates. In this context, I would like to underscore the envisioned potential for fostering improvisation and interplay through what I term "gesture interference" between performers – a 'hyper' jam session where mutual interaction and transformation of each other's musical material occur during a performance. This could manifest, for instance, in open sections of a score or within a completely improvised piece.

Finally, this research aims to contribute to resource sustainability by employing recyclable materials and developing strategies that promote the instrument's self-sufficiency in terms of energy consumption, a quality often lost with the introduction of electrified instruments.

Acknowledgments

This research is supported by Fundação para a Ciência e Tecnologia, I.P (FCT), grant 2022.11160.BD. I would like to thank my supervisors, Dr. Henrique Portovedo and Dr. Sara Carvalho for their contribution to this research, Dr. Vitor Sencadas from the Department of Materials and Ceramics at Aveiro University, Philippe Trovão (MaxMSP implementation) and Dr. Gilberto Bernardes and Dr. Luís Aly at INESC TEC.

6. REFERENCES

[1] C. Small, *Musicking: The Meanings of Performing and Listening,* Wesleyan: University Press, 1998

[2] D. Van Nort, "Instrumental Listening: Sonic Gesture as Design Principle," *Organized Sound*, 14(2), pp. 177–187, 2009.

[3] J. Malloch et al., "Towards a Conceptual Framework for Digital Musical Instruments," in *Proc. Of the 9th Int. Conf. on Digital Audio Effects (DAFx-06), Montreal, Canada, 2006*, pp. 49–52.

[4] J. d'Escriván, "To Sing the Body Electric: Instruments and Effort in the Performance of Electronic Music," *Contemporary Music Review*, Vol.25, pp. 183–191, 2006.

[5] S. Jordà, "Instruments and Players: Some Thoughts on Digital Lutherie," *Journal of New Music Research*, 33(3), pp. 321–341, 2004.

[6] R. Hamilton, "Mediated Musical Interactions in Virtual Environments," in S. Holland et al. (eds.) *New Directions in Music and Human-Computer Interaction* (pp. 243–257). Cham: Springer, 2019.

[7] T. Magnusson, "Designing Constraints: Composing and Performing with Digital Musical Systems", *Computer Music Journal*, 34(4), pp. 62–73, 2010.

[8] H. Portovedo et al., "HASGS: Five Years of Reduced Augmented Evolution", in *Proc. of the Conference on New Instruments for Musical Expression – NIME,* 2021. https://doi.org/10.21428/92fbeb44.643abd8c

[9] A. Tanaka, "Sensor-Based Musical Instruments and Interactive Music", in *Oxford Handbook of Computer Music,* pp. 233–257, 2009

[10] R. Fiebrink and L. Sonami, "Reflections on Eight Years of Instrument Creation with Machine Learning", in *Proc. of the International Conference on New Interfaces for Musical Expression – NIME,* 2020, pp. 237–242. https://doi.org/10.5281/zenodo.4813334

[11] J. Bullock and A. Momeni, "ml.lib: Robust, Cross-platform, Open-source Machine Learning for Max and Pure Data", in *Proc. of the International Conference on New Interfaces for Musical Expression – NIME'15*, 2015.

[12] B. Donato, J. Bullock, and A. Tanaka, "Myo Mapper: A Myo Armband to OSC mapper", in *Proc. of the International Conference on New Interfaces for Musical Expression – NIME'18,* 2018.

[13] C. Erdem and A. R. Jensenius, "RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation", in *Proc. of the International Conference on New Interfaces for Musical Expression – NIME'20*, Birmingham, UK, 2020.

[14] B. Di Donato, "Leap Mapper", *Github*. 2016 https://github.com/balandinodidonato/LeapMapper

[15] M. Wright and A. Freed, "Open Sound Control: A New Protocol for Communicating with Sound Synthesizers", in *Proc. of the International Computer Music Conference*, pp. 101–104, 1997.

[16] I. Hattwick and K. Umezaki, "Approaches to Interaction in a Digital Music Ensemble", in *Proc. of the International Conference on New Interfaces for Musical Expression* – NIME'12, Michigan, EUA, 2012.

[17] S. Carvalho, "Gesture as Metaphorical Process: An Exploration Through Musical Composition", *Revista Vórtex*, v.5, n.1, pp. 1–9, 2017.

[18] T. Wolf, "A Cognitive Model of Sight-Reading", *Journal of Psycholinguistic Research*, 5(2), pp. 143–171, 1976.

[19] L. Zattra, "Alvise Vidolin interview by Laura Zattra. The role of the computer music designers in composition and performance", in *Live-Electronic Music: Composition, Performance, Study*. Routledge pp. 83–101, 2018.

[20] K. Nymoen, M. R. Haugen, and A. R. Jensenius "MuMYO – Evaluating and Exploring the MYO Armband for Musical Interaction", in *Proc. of the International Conference on New Interfaces for Musical Expression* – NIME'15, Baton Rouge, EUA, 2015.

[21] J. Bullock et al., "Live Electronics in Practice: Approaches to training professional performers", *Organised Sound*, 18(2), pp. 170–177, 2013.

[22] J. Freeman, "Extreme Sight-Reading, Mediated Expression, and Audience Participation: Real-Time Music Notation in Live Performance", *Computer Music Journal*, 33(3), pp. 25–41, 2008.

[23] J. Winkel et al., "Battery-Free Game Boy", in *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3), Article 111, 2020.

[24] S. Umesh and S. Mittal, "A Survey of Techniques for Intermittent Computing", *Journal of Systems Architecture*, 11, 2021.

XNK AND HANDS2MIDICHANNELS: NEW SOFTWARE TOOLS FOR COMPOSERS AND IMPROVISERS

Carlos Mauro
CMG Solutions – Tufts University
cmg.solutions.a@gmail.com
carlos.mauro@tufts.edu

ABSTRACT

In the evolving landscape of 21st-century music, composers are turning to software tools to enhance creative workflows and improve composing efficiency. This paper introduces two novel tools independently developed by the author: Xnk and Hands2MIDIChannels. Xnk is a deterministic graphic music notation parser that processes .PNG images to produce a .TXT file suitable for the bach.roll object in MaxMSP. By merging qualities of abstract graphics and traditional notations, Xnk provides an alternative approach for composers in music sketching and composition. Meanwhile, Hands2MIDIChannels allocates MIDI events corresponding to a keyboard instrument to specific MIDI channels based on the hand that triggered the event. This allocation is derived by cross-referencing the original MIDI file with a synchronized video capturing the performer's hand movements over the keyboard. Notably, this tool offers invaluable utility to composers, improvisers, and musicians with disabilities. It enhances compositional efficiency by automating hand-based event assignation, enabling the swift transformation of improvisations into readable scores. The paper will explore the technical aspects, justification, current limitations, and potential avenues for both tools.

1. INTRODUCTION

Generalizing about the current state of contemporary music composition proves to be a hard task. Nevertheless, there is a consensus that technological advances in the last 25 years have had an unprecedented influence on all art forms. Nowadays, at the dawn of artificial intelligence, the previous statement resonates loudly.

In the case of concert music composition, technological advances in recent decades have been used by composers as a source of inspiration and as valuable tools. Despite the new workflows made possible by technological advances, there is still a lot of room for new technologies tailored for music composers and improvisers.

The first software discussed in this paper is Xnk, a deterministic graphic music notation parser. The name derives from the Romanian-born Greek-French composer, architect and engineer Iannis Xenakis whose graphic scores inspired the development of the software in question. The Xnk software provides its users the power to sketch/compose music by drawing, define instrumentation, support arbitrary divisions of the octave, have complete control over pitch range and length of the object, and, to output a formatted string ready to be handled by the Bach library [1] in the MaxMSP [2] environment for further processing.

The second software that will be discussed is Hands2MIDIChannels, a tool that routes MIDI events to specific channels based on the hand that triggered the event. This is done by cross-referencing the original MIDI file with a video that captured the performer's hand movements over the keyboard. Hands2MIDIChannels drastically decreases the time it takes to create usable scores from improvisation sessions on keyboard instruments.

This paper aims to present Xnk and Hands2MIDIChannels to the music notation technology community, discuss their technical aspects, reason to exist, their practical applications, current limitations, and their possible futures. Both tools are undergoing patent review with the United States Patent and Trademark Office (USPTO), emphasizing my commitment to legal protection.

2. XNK

2.1 Overview

Originally created in Python and later recreated in C++, Xnk takes an arbitrary number of .PNG images of the same size as inputs. For every inputted .PNG image, Xnk identifies events by iterating through all the rows looking for consecutive filled pixels. When an event is identified, Xnk stores the coordinates of the event onset and its duration. After iterating through every .PNG image and extracting the event information, a chronological sort takes place followed by a pitch average sort. Chronologically sorting the events' onsets allows them to be organized from first to last. The pitch average sort allows Xnk to organize the voices from highest register to lowest register.

After sorting the events and voices, Xnk performs pitch mapping using either a user defined range or, in its absence, a predefined one.

Once the events are correctly mapped on the pitch space (y-axis), a time-mapping is performed where the original onset and duration values (x-axis) of the events are scaled by a factor determined by user input. The pitch and time

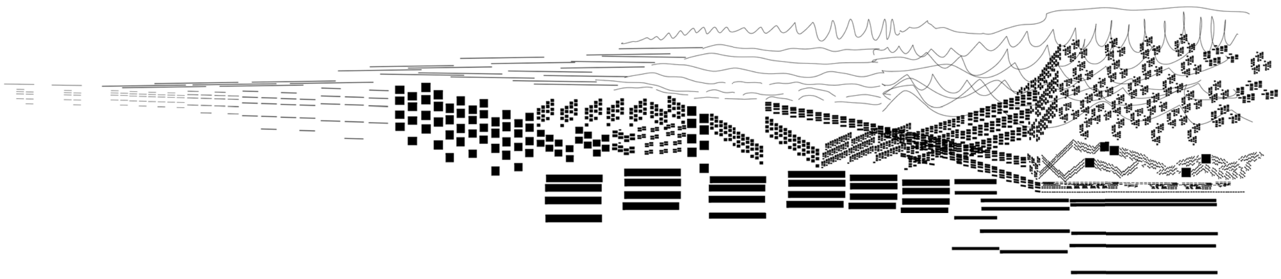


Figure 1. Superiomposed representation of .PNG layers.

mapping guarantees a deterministic and faithful graphic to traditional music notation translation.

At this point, the data is finally ready to be converted into a string using the Bach's library tree format. This allows the data from Xnk to be opened in MaxMSP and later through any mainstream music editor.

Link: Video of Xnk demonstration

2.2 Detailed Explanation

Xnk uses the following C++ libraries: OpenCV2, LibPNG.

2.2.1 Preparing the .PNG Images

Xnk interprets individual .PNG images as independent voices. This gives the composer the power to think about each image as either: individual instruments or an ensemble of instruments. Because Xnk preserves the independence of layers, each image will correspond to an individual voice in the output. If the user inputs N number of images, the output will have N number of voices.

The .PNG images can be created using tools such as Adobe Illustrator, Concepts, Inkscape, etc. When using these tools, the user needs to export each layer individually and store it as an isolated layer. This will ensure that every layer is of the same size. See Fig. 1.

2.2.2 User-Defined Parameters

As Fig. 2 shows, Xnk gives its users control over the following parameters:

- Maximum Output (highest) Pitch.
- Minimum Output (lowest) Pitch.
- MIDI Velocity.
- Target Length.

The parameters Maximum and Minimum Pitch define the pitch range of the output. These parameters expect the user to input values in midicents. If not defined, they default to a range from 10800 to 2100.

The MIDI Velocity parameter defines a uniform MIDI velocity for the output. It takes integer values ranging from 0 to 127.

The Target Length parameter is used to calculate the stretch factor that will be used for scaling the events' onsets and duration. This parameter expects the user to input the desire length of the output in milliseconds.

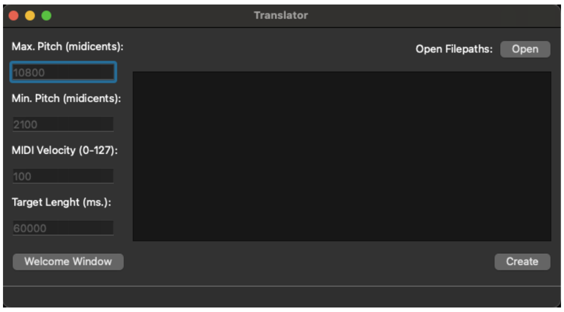


Figure 2. Xnk' Parameter Input.

2.2.3 Pre-Processing Preparation

Before any processing takes place, Xnk reads and stores the User-Defined Parameter values as well as the filepath list that contains the location of the .PNG images.

2.2.4 Image Processing

1. An empty vector is created that will be used as a container that will store the binarized images.
2. Xnk iterates through each string element in the filepath vector created on the previous step. For each filepath in the filepath vector:
    - Loads the image found at the iterated filepath to memory.
    - Binarization of the image. This is done so that Xnk only deals with 1s and 0s. This makes further processing simpler and more efficient.
    - Adds the binarized image data to the binarized image container vector created before the start of the iteration loop.
    - Cleans up memory preparing the next iteration.
3. Creates an empty vector that will store vectors containing the Voice Event Data. The nested vector structure is used to preserve the independence of each inputted .PNG image.
4. Once all the images are binarized and stored to memory, Xnk starts the event identification process by iterating through every binarized image stored at the binarized image container vector. For each binarized image in the binarized image container vector:
    - Initializes an empty vector that will serve as a local event container for the event data belonging the current binarized image iteration.

- Iterates through every row.
- If it detects a series of contiguous filled pixels with a length > 2, it identifies it as an event and records the X and Y coordinates of the event onset, along with its duration, in the local event container.
- Once a row iteration finishes, Xnk commits the Local Event vector container to the Voice Event Data vector container.
- Xnk cleans up the Local Event container and moves on the next row iteration.
- After iterating through all the rows, Xnk moves on to the next image in the binarized image vector.

2.2.5 Event and Voice Sorting

Once Xnk finishes extracting event information from the inputted .PNG images, it is time to sort the data. The first sort that takes place is a chronological sort.

To sort the data chronologically, Xnk iterates through every voice in the Voice Event Data Container vector. For every voice inside Voice Event Data Container:

- It organizes the events' onsets so that they are stored in chronological order.

After all the events inside the Voice Event Data Container vector are organized chronologically, a pitch average sort takes place.

To sort the Voices inside the Voice Event Data Container vector, Xnk iterates through every voice in the Voice Event Data Container vector. For every voice inside Voice Event Data Container:

- The pitch average (y-axis) of the event coordinates is calculated for each voice.

Once the average is calculated, Xnk reorganizes the data from highest pitch average to lowest pitch average. This is done so to improve the output readability.

2.2.6 Pitch and Time Mapping

After organizing the data appropriately, Xnk proceeds with Pitch and Time Mapping processing. To map the y-axis component of events to a pitch, Xnk identifies the maximum (I_{mx}) and minimum (I_{mn}) y-axis values from the events extracted from the inputted .PNG images. Additionally, Xnk consults the User Defined Parameters Maximum Output Pitch (O_{mx}) and Minimum Output Pitch (O_{mn}) to define the output pitch range.

With this information, Xnk computes the pitch mapping using the following function:

$$P = (O_{mn} + O_{mx}) - \left[(Y - I_{mn})\left(\frac{O_{mx} - O_{mn}}{I_{mx} - I_{mn}}\right)\right] + O_{mn} \quad (1)$$

Where:

- P = Mapped Pitch.
- Y = y-axis component to be mapped.
- O_{mx} = User defined output Maximum Pitch.
- O_{mn} = User defined output Minimum Pitch.
- I_{mx} = Maximum y-axis value from input data.
- I_{mn} = Minimum y-axis value from input data.

After mapping all the events belonging to all voices inside the Voice Event Data Container vector, Xnk scales the events onsets and durations to reach the User Defined Target Duration.

To do this, Xnk needs to find the Time Scaling Factor (*TSF*). To this end, Xnk finds the and maximum x-axis value (T_{mx}) present on the Voice Event Data Container vector. Additionally, Xnk refers to the User Defined Parameter Target Length (*TL*).

The *TSF* is calculated using the following equation:

$$TSF = Tl/T_{mx} \quad (2)$$

Where:

- *TSF* = Time Scaling Factor.
- *Tl* = Target Length.
- T_{mx} = Maximum x-axis Value.

Xnk then scales all the events' onsets and durations located at the Voice Event Data Container vector by the previously calculated TSF.

2.2.7 Data to Bach's llll Conversion

The final processing step of Xnk is making the processed data readable by the Bachroll object in MaxMSP. To do this, a .TXT file containing a string in MaxMSP's Bach's library llll format is created. The overarching llll format is as follows:

$$[[< voice1 >][< voice2 >][< voice3 >]]$$

Where each $voice\#$ is defined as:

[o [p d v]] [o [p d v]] ...

Where:

- *o* = Event Onset.
- *p* = Event Pitch.
- *d* = Event Duration.
- *v* = Event MIDI Velocity.

Once Xnk iterates through all the voices and the events inside the Voice Event Data Container vector, it creates the .TXT file containing the formatted string.

2.2.8 Opening in MaxMSP

The created .TXT can be opened inside MaxMSP by calling the read method of the bach.roll object. See Fig. 3

2.3 Justification

The purpose of Xnk is to provide its users with a tool that can translate graphical scores into music notation that supports an unlimited number of voices and arbitrary divisions of the octave. It provides its users a faithful representation for any inquired graphic score. Xnk strives to achieve this by making as few assumptions as possible.

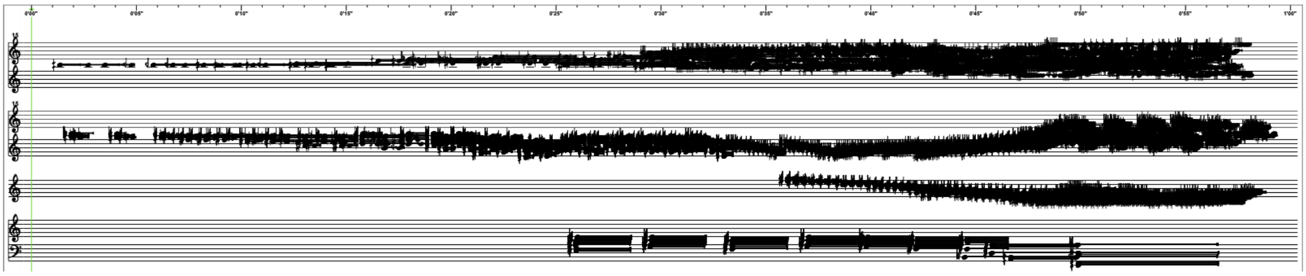


Figure 3. Output in MaxMSP.

While programs such as Finale or Sibelius are powerful tools, they are focused on editing music notation, not composing music. To do the latter, there are alternatives such as the Bach Library [1] for MaxMSP [2], OpenMusic [3], MaxScore [4], SPEAR [5], among others. Additionally, a similar concept was adopted by Hyperscore [6].

While SPEAR also deals with graphics to create sonorities, it is centered on spectral analysis of pre-existing sound files. In Hyperscore's case, it gave its users visual representation of music parameters to create music. Arguably, Hyperscore interpreted graphical notation. In contrast, Xnk parses graphical notation.

2.4 Current Limitations

At the current stage of development, Xnk has the following limitations:

- Only accepts .PNG images as inputs.
- MIDI Velocity is arbitrarily decided, not deducted
- The dependance on external graphic editors to create scores and parts does not allow Xnk to natively do graphic edits to the .PNG images. Furthermore, Xnk's relies on MaxMSP for further processing, playback, and MusicXML/MIDI export methods.
- The algorithm that extracts event data from binarized images needs further improvement.
- Not able to recognize glissandos as individual events. They are interpreted as sequences of events.
- Extreme sensitivity creates a substantial amount of redundant data.

2.5 Future Development

Currently, there are four main priorities for future development. In order of importance, they are:

- Create a native graphic notation editor.
- Improve the event extraction algorithm.
- Create a native playback engine.
- Create a native music notation editor.

2.5.1 Native Graphic Notation Editor

Developing a Native Graphic Notation Editor will be a huge step forward in Xnk's development. This will enhance the Xnk's workflow by eliminating the need to constantly alternate between Xnk and the user's graphic editor of his choice. Rather, the user will be able to make edits directly in the Xnk environment.

Developing this feature will prepare the path for in-depth user defined MIDI velocity handling as well as Computer Assisted composition AI algorithms.

In its current state, Xnk has a graphic editor that is at an early stage of development. The current challenge with the editor environment within Xnk is that it needs to be able to expand horizontally towards positive x infinitely while, at the same time, supporting infinite zoom. This is currently being done by treating graphics as vectors in a similar way in which Adobe's Illustrator [7] or TopHatch's Concepts [8] do it.

Once the previously discussed features of the editor environment are developed, specific graphic editing in the context of Xnk's goals will be created.

2.5.2 Native Playback Engine

Playback engines are indispensable for music software. As mentioned before, Xnk currently relies on MaxMSP to handle playback. Developing a Native Playback Engine will improve Xnk's functionality, practicality and workflow.

This playback engine must be powerful, scalable, polyphonic, and microtonal friendly in nature. One of the current questions is whether to use MIDI.

The MIDI protocol was not designed to natively work with microtones. While there are ways around that, they can be unpractical in their application and decrease the overall power of the playback engine. The most appealing solution would be to create a native Xnk's playback protocol that would allow for polyphonic microtonality without relying on workarounds. In addition to being able to handle microtonality natively, this theoretical new protocol must be able to be translated into commonplace formats and protocols such as MIDI or MusicXml.

2.5.3 Native Music Notation Editor

The creation of a Native Music Notation Editor will make Xnk an powerful and versatile music composition program.

One of the most important and challenging aspects of developing this music editor will be implementing the logic behind rhythmic quantizing tools.

Contrasting Xnk editor with the ones present in mainstream music editors, Xnk will be composition focused

Figure 4. Hands2MIDIChannels raw input video.

rather than engraving focused. The goal of Xnk is to provide its users with a flexible and dynamic composition environment.

2.6 Closing Thoughts

Xnk is not intended to replace industry-standard music editors. In the case of Finale [9] and Sibelius [10], they are both powerful computer engraving programs. However, they lack the degree of flexibility and functionality needed for a pure composition software.

Ideally, Xnk will become an intermediate step between music sketching and a finalized score.

3. HANDS2MIDICHANNELS

3.1 Overview

Hands2MIDIChannels is a python script capable of routing MIDI events to specific channels based on the hand that triggered the event. This is done by cross-referencing the original MIDI file with a video that captured the performer's hand movements over the keyboard.

It takes a MIDI file and a corresponding video as inputs. After the video file passes through a number of pre-processing steps, it is manually synchronized with the video recording. This is done so that Hands2MIDIChannels knows the specific moment in which a MIDI event is triggered in the corresponding video.

At this point, further pre-processing steps take place in preparation for the cross-reference stage. During the cross-reference, Hands2MIDIChannels iterates through the events on the MIDI file as shown in the video and calculates which hand most likely triggered the iterated event.

After Hands2MIDIChannels assigns a hand to every relevant event in the MIDI file, it creates a new MIDI file where the events are routed to specific channels based on the hand that triggered the event.

Link: Video of Hands2MIDIChannels demonstration

3.2 Detailed Explanation

Hands2MIDIChannels uses the following Python libraries: OpenCV2, Numpy, Mediapipe, MIDO, Pillow, and TQDM.

3.2.1 Initial Video Preparation

The goal of the Initial Video Preparation stage is to make the video recording as computer-vision analysis-friendly as possible. At the end of this stage, Hands2MIDIChannels obtains a video where the area of interest shows the keyboard as a perfect rectangle. This makes further processing straightforward.

1. Store Input Paths
    - Hands2MIDIChannels stores the paths for the MIDI and the corresponding video file.
2. Initial Video Crop
    - User defines the overall area of interest that clearly shows the performer's hands as well as the keyboard using an interactive interface. See Fig. 4 and 5.
3. Set Keyboard Top Bound
    - User defines the keyboard top bound using an interactive interface.

Figure 5. Hands2MIDIChannels input cropped video focusing on the area of interest.

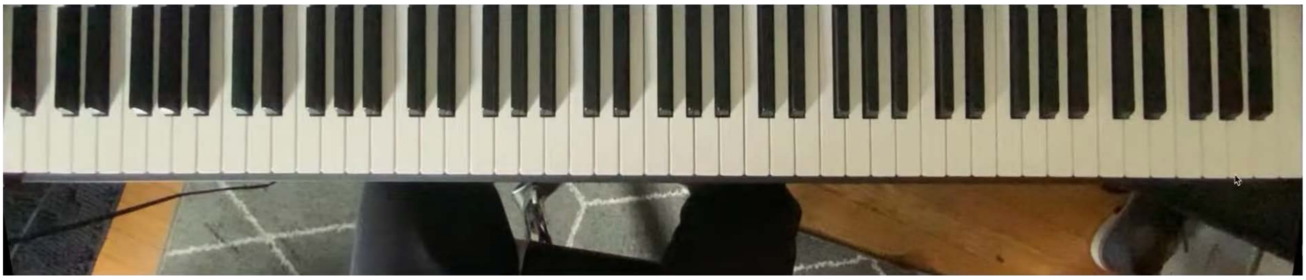


Figure 6. Hands2MIDIChannels input video after pre-processing and transformation.

4. Set Keyboard Bottom Bound
    - User defines the keyboard bottom bound using an interactive interface.
5. Keyboard Bound Calculation
    - Hands2MIDIChannels uses the previously gathered information to calculate the side bounds of the keyboard.
6. Hand Bounds Calculation
    - Hands2MIDIChannels calculates the hand bounds in respect to the keyboard bounds.
7. Calculate 2D Transformation Matrix
    - To account for image distortion commonly created by lenses, Hands2MIDIChannels calculates a 2-dimensional transformation matrix that will make the keyboard top and bottom bounds be as parallel as possible over the x-axis. See Fig. 6.
8. Set Y Levels for Black and White Keys
    - User defines the bottom bounds for both: the black and white keys. Note that thanks to the previous 2D Transformational Matrix, this bound is parallel to the x-axis.

3.2.2 Video-MIDI Synchronization

Synchronizing video and MIDI as well as abstracting the individual onsets for each MIDI event is important because this will help Hands2MIDIChannels locate the exact place in time on the video where an inquired MIDI event takes place.

1. Set Sync Onsets
    - User inputs the first and the last onset of events in the MIDI file as they are shown in the video.
2. Individual Onset Calculation
    - By using the first and last onset, Hands2MIDIChannels calculates the individual place in time where a MIDI event is shown at the video.

3.2.3 Final Preparations

During the Final Preparations Stage, Hands2MIDIChannels calculates the contours corresponding to every piano key. Then, the user is presented with an interactive interface that allows finetuning the contours. These contours will be used in the Cross-referencing Stage that follows.

1. Hands2MIDIChannels creates an approximation of the contours corresponding to every individual white and black keys.
2. The user adjusts the contours so that they accurately represent the keyboard as shown in the video.
    - Because the video is transformed by a 2-Dimensional Matrix, the keys in the keyboard tend to have slightly different widths. Modifying the contours' placement over the x-axis accounts and corrects for this.
3. The final keyboard' keys contours are created. MIDI note information is attached.

3.2.4 Cross-Referencing

During the cross-referencing stage, Hands2MIDIChannels iterates through the MIDI events and their corresponding onsets in the video. It calculates which hand most likely triggered the iterated event. See Fig. 7. For each MIDI Event:

- Get MIDI note number.

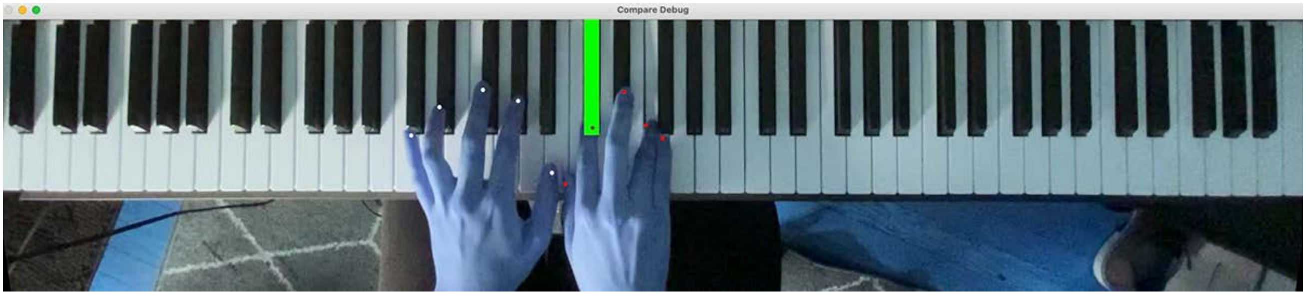


Figure 7. Hands2MIDIChannels recognizing hand data from a still frame. Inquired MIDI event is highlighted in green.

- Get video onset.
- Get keyboard key contour corresponding to the inquired MIDI note number.
- Load the inquired video onset as an image.
- Calculate which hand most likely triggered the event.
- Store result.

Hands2MIDIChannels uses logical gates to determine if a note was played with the right or left hand. Currently there are 4 stages in determining the guess. If the algorithm arrives to a guess at any stage, it skips the remaining stages.

The stages are:

- Check if no hands are detected in the video. In this case, Hands2MIDIChannels routes the event to an excluded MIDI channel. This stage is often triggered by a low frame rate video input or by Mediapipe failing to recognize hand information.
- Check if there is only one hand detected in the video. In this case, Hands2MIDIChannels assigns the detected hand as the hand that triggered the inquired MIDI event.
- Check if there is a hand located inside the inquired key contour area (that was calculated in the Final Preparations stage of the program). If this case is True, the guess will be the hand that is inside the inquired key area.
- Calculates the hand closest the inquired key area. This is the last resource. In this case, the guess is the hand closest to the inquired key area.

3.2.5 Create Output

In the Create Output stage, Hands2MIDIChannels creates a new MIDI file where the events are routed to specific channels based on the hand that triggered the event. This is done by duplicating the original MIDI file and updating the channel attribute for the MIDI events. See Fig. 8.

3.3 Justification

Creating a legible piano score from a MIDI recording is a laborious task. This is because most DAWs and music editors record MIDI events to a single channel by default. This in turn routes all events to a single music staff. To fix this, there are two common solutions that industry standard software present: manual channel assignation or arbitrary choice of split point.

The manual channel assignation method involves the user changing the MIDI channel for each MIDI event manually. This can be a practical solution for short recordings but it can quickly become an enormous task when dealing with larger durations.

The second method, arbitrary choice of split point, involves the user selecting a note as the split point. The notes below the split point will be routed to an individual MIDI channel while the notes above the split point will be routed to another MIDI channel. For simple purposes, this works great. However, for contemporary music and improvisations, this method does not work since there isn't a constant split point most of the time.

Hands2MIDIChannels offers an alternate solution that strives to involve minimal user input and improve efficiency. At the same time, Hands2MIDIChannels opens the door for accessibility capabilities. Blind improvisers and composers will be particularly benefited by this program since it drastically decreases the need for external intervention.

3.4 Current Limitations

Hands2MIDIChannels aims to decrease user involvement, however, it still relies heavily on it. Additionally, processing speed has room for improvement. While it remains faster than any other method, processing times for long files (28000+ events) is around 2:30:00 (hr:mn:sc). Furthermore, Hands2MIDIChannels only works with MIDI keyboards. It does not support acoustic instruments.

Finally, the guess accuracy has room for improvement. Errors created by Mediapipe not recognizing hand landmark information or a slight synchronization error in the video create inaccurate guesses.

3.5 Future Developments

Future developments for Hands2MIDIChannels aim to achieve the following:

1. Improve the hand-guess algorithm.
2. Support acoustic pianos recordings.
3. Improve time efficiency.
4. Decrease user intervention as much as possible.

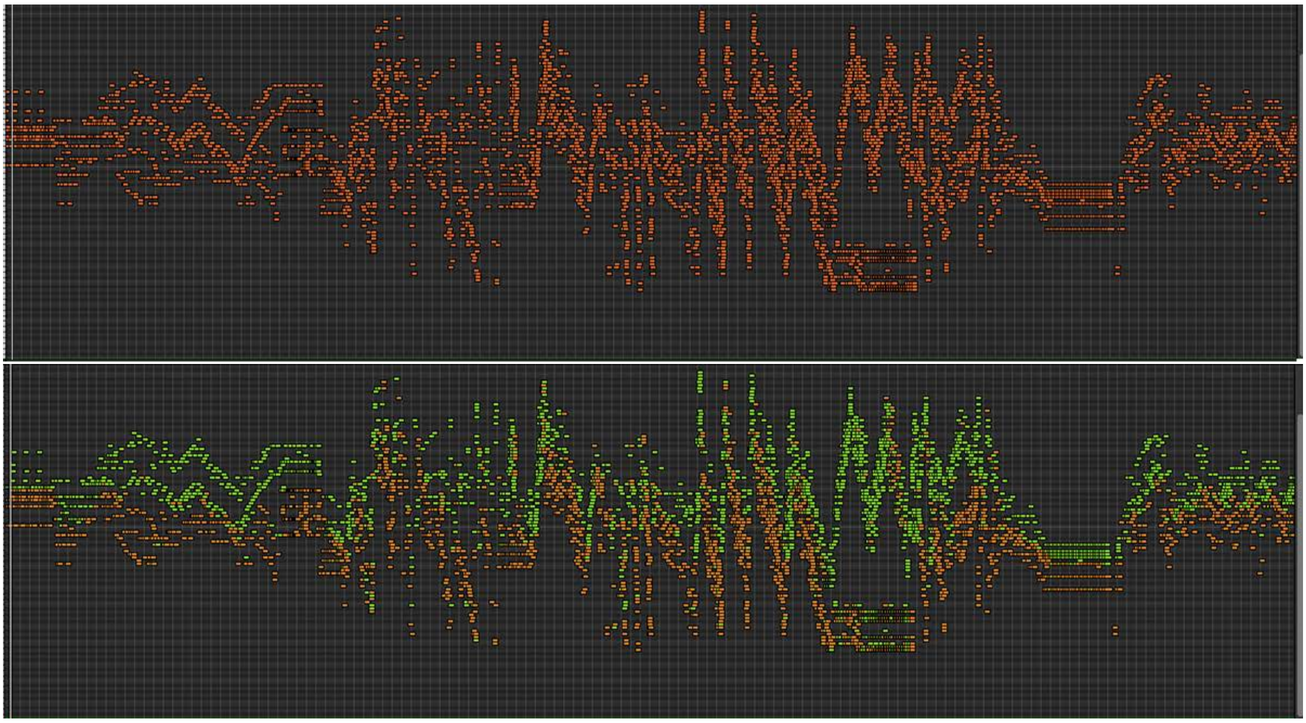


Figure 8. Before and after Hands2MIDIChannels processing comparison. MIDI channels are represented by colors.

To improve the hand-guess algorithm in the future, it would be useful to create a machine learning model from scratch that specializes in recognizing hand information from a frame. Another possible option would be to add additional steps in the video pre-process stage with the goal of making the visual material as compatible as possible with Mediapipe. Yet another possible path is to implement a function that will audit the guesses and calculate their accuracy. Based on this probability, the function will be able to correct the guesses if they need it.

Supporting acoustic pianos video recordings as a singular input for Hands2MIDIChannels is a challenge, yet it is a theoretically possible challenge. By abstracting pitch information from the recorded signal using Fast Fourier transformations, comparing it to the corresponding video frame, then having a machine learning model recognize the hand data in that frame, and finally, output an accurate MIDI file with events sorted in two channels, Hands2MIDIChannels would be able to support acoustic piano video recordings.

To improve time efficiency, the code needs to be reorganized, translated to C++, and designed to decrease user involvement as much as possible. Hands2MIDIChannels aims to be a efficient tool that is able to save its users' time by automating the hand assignation process as much as possible

3.6 Closing Thoughts

Hands2MIDIChannels gives its users a time-efficient tool that allows them to improvise/compose freely at the keyboard with the certainty that they will have access to a organized MIDI file ready for time quantization. Because of its nature, Hands2MIDIChannels thrives as both: as a standalone software, and as an external function capable of interacting with existing DAWs and music editors.

Finally, Hands2MIDIChannels is of particular help for blind composers and improvisers.

4. CONCLUSION

This paper introduces two novel software tools independently developed by the author, offering innovative solutions for optimizing creative workflows in music composition and improvisation.

Addressing specific challenges overlooked by current industry-standard software, Xnk presents a deterministic, reliable, and scalable framework for the translation of graphic elements into traditional music notation. Conversely, Hands 2MIDIChannels streamlines the hand assignation process for keyboard instrument recordings.

Beyond their technical advantages, both tools contribute to unparalleled accessibility solutions, catering to individuals with disabilities. Moreover, they play a pivotal role in democratizing the music creation process, extending its reach to a broader and more diverse population.

In summary, these tools not only advance the field of digital music notation and transcription but also have the potential to significantly impact accessibility and inclusivity within the realm of music creation, marking a noteworthy stride towards a more universally accessible and democratized musical landscape.

Acknowledgments

I would like to extend my heartfelt thanks to Stella Gitelman Willoughby, whose unwavering support and insightful advice have been instrumental throughout the course of this project. Special appreciation is also due to my MA advisor, John McDonald, whose artistry, encouragement, and deeply inspiring conversations are shaping my academic journey at Tufts.

Furthermore, I am grateful to Melinda Latour, PhD, for her meticulous proofreading of the paper, enhancing its clarity and precision.

Lastly, I wish to express my sincere appreciation to the Tufts Music Department Graduate program for their invaluable financial assistance.

5. REFERENCES

[1] A. Agostini and D. Ghisi, "A Max Library for Musical Notation and Computer-Aided Composition," *Computer music journal*, vol. 39, no. 2, pp. 11–27, 2015.

[2] Cycling'74, "Max." [Online]. Available: https://cycling74.com/products/max

[3] G. Assayag, C. Rueda, M. Laurson, C. Agon, and O. Delerue, "Computer-Assisted Composition at IRCAM: From PatchWork to OpenMusic," *Computer Music Journal*, vol. 23, no. 3, pp. 59–72, 1999.

[4] G. Hajdu and N. Didkovsky, "MaxScore: Recent Developments," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'18*, S. Bhagwati and J. Bresson, Eds. Montreal, Canada: Concordia University, 2018, pp. 138–146.

[5] M. Klingbeil, "Software for spectral analysis, editing, and synthesis," in *International Computer Music Association*, vol. 2005, 2005.

[6] M. M. Farbood and E. Pasztor, "Hyperscore graphics software package," 2004.

[7] Adobe Inc., "Adobe illustrator." [Online]. Available: https://adobe.com/products/illustrator

[8] Tophatch Inc., "Concepts." [Online]. Available: https://concepts.app/en/

[9] MakeMusic Inc., "Finale." [Online]. Available: https://www.finalemusic.com/

[10] Avid Technology Inc., "Sibelius." [Online]. Available: https://www.avid.com/sibelius

SKETCHING MAGNETIC INTERACTIONS FOR NEURAL SYNTHESIS

Nicola Privato*
Intelligent Instruments Lab
Reykjkavík, Iceland
nicola@lhi.is

Giacomo Lepri*
Intelligent Instruments Lab
Reykjkavík, Iceland
giacomo@lhi.is

Thor Magnusson
Intelligent Instruments Lab
Reykjkavík, Iceland
thor.magnusson@lhi.is

Einar Torfi Einarsson
Iceland University of the Arts
Reykjkavík, Iceland
et@lhi.is

ABSTRACT

The application of neural synthesis for sound generation has grown significantly in recent years. Models such as RAVE offer real-time control by mapping sounds to and from numerical vectors in an abstract latent space whose features of entanglement, arbitrariness, and continuity pose novel challenges to musicians and composers in encoding and interpreting the inscription. In this paper, we introduce Stacco, a magnetic score system that addresses these challenges by functioning as a performative and compositional support for neural synthesis models through embodied sketching. We describe the system and present insights from a workshop aimed at exploring the compositional potential of this platform. We conclude by reflecting on how, during the workshop, Stacco's playfulness and magnetic materiality translated into the participants' scores and on the broader implications of embodied sketching in notating music for neural synthesis.

1. INTRODUCTION

Over the past few years there has been notable advancement in neural synthesis techniques for music generation, with models such as RAVE [1] offering real-time control and the direct navigation of their latent representations.

As neural audio synthesis for real-time applications is a very recent technology, little work has been done in developing specific performative and compositional strategies for such tools, whose inherent features of parameter entanglement, arbitrariness, and continuity pose new challenges for composers and performers working with digital audio.

Within this domain, and more generally in sonic interaction design, one of the risks is to focus on the abstract layers characterizing software models, where the development of the interface often prioritizes technical and sonic factors, with little attention to the bodily experience of the performer [2].

Based on these premises, the contribution of this paper is twofold. On the one hand, we introduce Stacco, an *instrument-score* based on magnetic sensing that functions as an ideal performative and notational mediator for neural synthesis models, thus enabling us to reflect upon the unique compositional challenges that this novel synthesis technology brings forth. On the other hand, we address these challenges by reporting on the outcomes of a workshop with Stacco, where participants creatively notated musical scores on the body of the instrument as they experimented with a series of neural synthesis models.

In the next section, we discuss relevant literature and existing practices to situate our work, with a focus on the overlapping domains of notation and instrument design.

2. BACKGROUND

In the past few decades the field of *Digital Musical Instruments* (DMI) has witnessed a flourishing of innovative interdisciplinary methodologies, as musicians, designers and technologists have integrated diverse fields into their research practice. This concoction of viewpoints, ranging from media studies [3], social sciences [4] and political studies [5] represents a shift, and at times a dissolution, of the boundaries that traditionally separate musical categories and roles.

Gurevich describes a practice-based attitude to instrument design and performance which promotes an "ecological view of music-making" challenging the traditional model of music as communication, the rigid distinction between composer-performer-listener as well as the locus of agency across human and non-human factors [6]. Media theorist and sociologist Jonathan Sterne argues that "[i]n music technologies, media collapse into instruments – or, rather, the line between instruments and media grows fuzzy" [7]. Similarly, Battier and Schnell introduce the term *composed instrument* to underlay the fact that "computer systems used in musical performance carry as much the notion of an instrument as that of a score" [8].

Our work resonates with such attitudes and perspectives, and, more specifically, it builds on the notion of *instrument-score*, bringing into the debate new method-

*Nicola Privato and Giacomo Lepri contributed equally to the development of Stacco. Nicola Privato led the activities, collected and elaborated the data used in this paper.

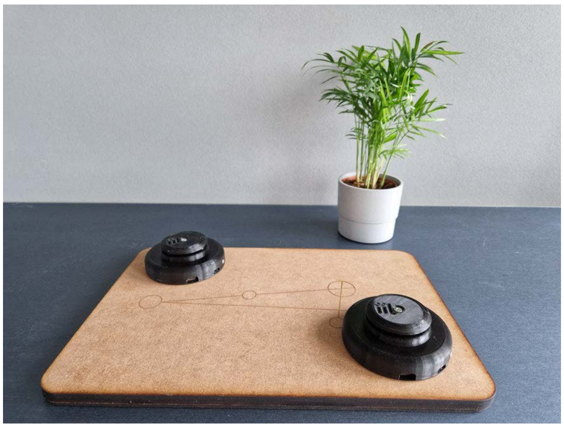


Figure 1. Nicola Privato's Thales.

Figure 2. Giacomo Lepri's Chowndolo.

ological and conceptual tools. Of particular relevance is here the work of Tomás and Kaltenbrunner who, drawing on Alvin Lucier's interviews [9], introduce the concept of *inherent scores* to describe the progressive embedding of inscriptions within the instrument [10] itself. Building on this, Tomás develops his *tangible scores:* a particular kind of inherent score allowing sound production through the tactile interaction with drawings engraved on a surface [11].

2.1 Magnetic Scores

Building upon Tomás' contribution, we describe Stacco, the musical interface introduced in this paper, as a *magnetic score* [12]: a particular type of inherent score featuring embedded magnets and a visual interpretation of their magnetic fields. To perform with a magnetic score means to interact with its magnetic fields, usually by means of other magnetic elements, as the variation in the magnetic field is detected and converted into sound.

One notable example of magnetic scores is Thales [13] (Fig. 1), which combines an engraved board embedding a series of magnets with two controllers allowing the tangible navigation of the magnetic fields. As the musician moves the controllers on top of a score with embedded magnets, the magnets in the controllers and those in the score repel or attract each other, suggesting various performative gestures.

Another relevant artwork is the Chowndolo [14] (Fig. 2), a magnetic pendulum whose trajectories are altered by magnets placed underneath the device. The chaotic oscillations of the pendulum are then translated into sound. The magnets below the pendulum can be arranged to compose new shapes: different configurations will produce unstable sonic variations, articulating music that evolves based on the pendulum's dance.

The instrument-scores mentioned above have a distinct embodied character, as they encode the inscription within the materiality of the interface, with the performer interpreting it through touch and gestures. This materiality is apparent in the score's resistance and action in response to the performer's gestures, empowering creativity and boosting the communication between the performer and the audience [15]. Our research aims to leverage such material features, exploring embodied ways to embed articulated compositional ideas, and how this might help composing and performing with neural synthesis.

2.2 Embodied Sonic Sketching

The investigation of embodied musical interactions encompasses different fields, including digital instrument design, sound design, dance, and theatre performance. Drawing on *soma design* [16], Avila et al. describe a set of workshops for the design of digital musical interfaces emerging from bodily explorations of guitar performance [2]. Delle Monache and Rocchesso thoroughly investigate embodied approaches to sound generation and manipulation and propose a set of interactive tools to sketch sounds using the voice [17, 18]. Based on long-lasting collaborations with dancers and musicians, Camurri et al. develop computational tools and frameworks to extract and interpret meaningful gestural information to be exploited while designing interactive experiences [19, 20].

Particularly relevant for our methodology is Kristina Andersen's *Magic Machines* workshop [21]. Drawing on diverse performance and theatre practices [22], Andersen proposes a playful design fiction approach to technology ideation, which involves the creation of mock-up objects that work *as if by magic*. Andersen exploits the notion of the "magical unknown" to free a participant's imagination and generate manifestations of original technologies and interactions. Crucial to the Magic Machine workshop is the notion of "thinking with the hands" [23], where the physical making and manipulation of artefacts allows participants to sketch ideas through the spontaneous and intuitive assemblage of materials. Lepri and McPherson have also exploited Andersen's Magic Machines workshop to facilitate the emergence of subjective aesthetic values and priorities of artists engaged in the design of novel musical interactions [4].

The activity presented in this paper may be viewed as a variation of such work, as participants were invited to augment Stacco with mundane objects, by sketching physical scores to be placed on the instrument itself and guiding the performance with an interface controlling a neural synthesis engine.

2.3 Neural Synthesis

The recent introduction of deep learning methods has brought forth exciting technologies for the generation of raw audio [24, 25]. Whereas the slow responsiveness of neural audio synthesis models has initially limited their use in performative scenarios, the introduction of RAVE [1], a Real-time Audio Variational Autoencoder (VAE) that performs fast and high-quality audio synthesis, has drastically facilitated the use of neural synthesis in interactive contexts [13].

RAVE is a fine-grained latent variable model. It encodes a stream of raw audio to a stream of latent vectors, before decoding it back to audio. It combines a variational autoencoder with an adversarial reconstruction term, where realistic details are sampled by the decoder and the coarse qualities of the sound are represented in the *latent space:* a multi-dimensional, compressed representation, in between the model's encoding and decoding functions.

RAVE may be used for style transfer, by forwarding raw audio to its encoding function or explored by navigating its latent representation through control signals. It is also possible to combine these two approaches, for instance by forwarding audio through the encoding phase to one or two latent dimensions whilst driving the others with dedicated signals, or by mixing audio and control signals on individual latent dimensions. In the study described in this paper, we apply the former method, mapping each of Stacco's data points to a single latent dimension, and bypassing the encoding function.

3. STACCO

In this section, we provide a high-level analysis of neural audio synthesis' affordances from the perspective of the composer, motivating the reasons that led us to develop our framework for embodied sketching. We continue by describing Stacco's hardware and software, and the neural synthesis models we used in our study.

3.1 Composing for Neural Synthesis

Neural Synthesis is a new, exciting synthesis method whose distinctive features open new compositional approaches and expressive possibilities. In composing with RAVE, the artist curates datasets of raw sounds. During the training phase, the model learns to reconstruct the datasets and distributes its most meaningful sound features in the latent space.

In our artistic explorations of different models' latent space, we encountered three distinctive qualities that influence the interaction with the model and that, ultimately, redefine the compositional strategies:

- Entanglement: Latent dimensions are deeply intertwined; they appear as a complex and interdependent structure of relations, where any change in the state of one latent variable affects the behaviour of all the others. Consequentially, it is not possible to manipulate one single parameter, such as amplitude or frequency, without affecting the others.
- Continuity: Latent spaces represent continuous dimensions that remain active at all times and cannot be selectively deactivated. Consequently, systems employing RAVE are typically designed to offer continuous values rather than booleans.
- Arbitrariness: The distribution of the sound features within the latent spaces is autonomously performed by the model during the training phase, it differs with every dataset and initialisation seed, and is empirically explored *a posteriori* by interacting with the trained model. This implies that it is not possible to abstract gestural constants affecting the sound in similar ways with two different models.

Because of the aforementioned arbitrariness, composers need to spend time exploring every new model, looking for patterns, configurations, and gestures that lead to satisfactory acoustic results. This process is necessarily mediated by a physical interface, that influences the control of the system as well as the understanding of the algorithm [26].

Once the model has been thoroughly explored, and meaningful gestures and interesting areas within the latent space have been detected, the composer faces the challenge of defining an appropriate notational strategy. This is of course dependent on the overall aesthetic aim of the composer. For the case described in this paper, i.e. the navigation of RAVE through control signals, it is nevertheless possible to outline two general considerations.

First, because the latent distribution of a model is learned through a gestural exploration on a given interface, an immediate and intuitive approach to notation is the one that allows the reconstruction of those gestures on that interface; second, because sound features such as frequency or amplitude cannot be disentangled from the latent representation, it is not possible to notate them individually. This further binds the representation to the synthetic potential of the gesture.

As a consequence of the notation's gestural reliance, the act of composing becomes intrinsically intertwined with the interface. By building upon the capability of magnetic scores to incorporate the notation into the instrument, our instrument-score allows artists to compose by sketching gestures on interchangeable musical scores placed on the interface itself. In addition, Stacco features a one-to-one, unlabeled and continuous mapping to latent dimensions, and the interaction of the magnets, their agency and resistance makes the playing engaging and enjoyable.

3.2 Hardware

Stacco (Fig. 3) is a novel type of magnetic score embedding magnetic attractors and sensors underneath an engraved surface, and whose design features are aimed at facilitating the interaction and composition with RAVE whilst providing a rich and playful musical experience.

Stacco consists of four miniaturised magnetic discs [12] combining a magnetometer with a permanent magnet each. Each sensor performs two-dimensional readings of nearby magnetic fields whilst actively attracting and repelling nearby magnets and ferromagnetic objects. The sensors

Figure 3. Stacco.

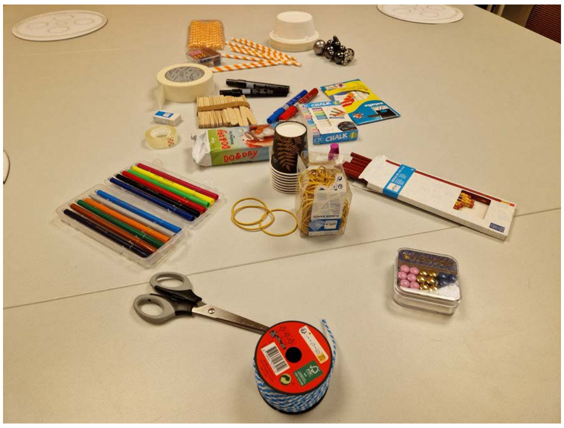


Figure 4. Workshop Material.

are displaced in four symmetrical points underneath an oval wooden board and connected to a Bela [27] for embedded synthesis. The 32 x 217 cm engraved board features a raised edge and is enclosed via a living hinge structure.

The performer interacts with Stacco by throwing and displacing on the board a series of magnetic spheres of variable dimensions. The four magnets under the board, each coupled to a magnetometer, actively engage the performer in a playful dance of agencies, enacting the inscription through the interactions of their magnetic fields with the spheres controlled by the performer.

The engraved board reveals the position of the four sensors and the magnetic field through a series of circular patterns. An additional smaller circle is engraved in the centre and may host magnets with opposed polarity for more complex interactions. A larger circle encloses all four attractors within the larger oval.

Stacco's design allows embedding sketches as tailored oval sheets with two or three-dimensional inscriptions. Since the spheres interact with the instrument through their magnetic field and do not need to touch the board, the presence of the oval sheets does not affect the instrument's playability.

3.3 Software

Stacco uses Bela [27] to forward the sensors' data to the laptop via OSC, for a total of eight data points, two axes per sensor.

We trained three RAVE models based on the following datasets, mapping each latent dimension to one of the eight data points forwarded by Stacco's magnetometers.

- Choir: model trained by the artist Jonathan Reus. The model is RAVE v1, with 16 latent dimensions mapped two-to-one with each of Stacco's sensor readings.
- Organ: recordings of open-source organ music. Small amounts of voice and other instruments are included and vinyl record noises are prominent. The model is RAVE v1 modified by Victor Shepardson, with 16 latent dimensions mapped two-to-one with each of Stacco's sensor readings.
- Magnets: one-hour recording of magnets of different dimensions interacting with each other or scratching wooden and metallic surfaces. The model is RAVE v1, 48Khz, 8 latent dimensions linearly mapped one-to-one with each of Stacco's sensor readings.

In addition to this, to compare neural synthesis with traditional synthesis methods we developed a stereo FM synth with eight control parameters, individually mapped to each of Stacco's sensor readings.

4. WORKSHOP

We explored the composed nature of Stacco in a workshop in Łodz (Poland) at Act In Out, a joint initiative by the Art Factory in Łódź, Slaturhusid Art Center (Iceland), Carte Blanche Dance Theatre and Visjoner Theatre (Norway). In this workshop, titled *Composing Magnetic Interactions*, we used Stacco as a support to the design of a series of musical scores, aiming to understand the compositional practices that may develop around neural synthesis and whether the concept of instrument-score might favour an embodied understanding of the model.

4.1 Description

The workshop involved five participants: three trained musicians (P1 to P3), one graphic designer (P4), and one visual artist and art curator (P5) (Fig. 4).

Before the event, we cut a series of cardboard ovals with a stylized version of Stacco's engraving printed on one of the sides. Thanks to the instrument's raised edge, the ovals fit the top of the board working as removable blank canvases for the participants' compositional ideas. We provided each of the participants with one oval, and the whole group with the following materials: paper straws and cups, magnets of different sizes and colours, pencils, pens, markers, clay, rubber bands, rope, scissors, chalks, wooden sticks, tape, wall gum.

In choosing the materials we avoided tool-kits, electrical components, sensors and software units as well as materials with distinct acoustic properties (e.g. boxes and wood), in order not to limit people's imagination and inventiveness. Following Andersen's approach, we privi-

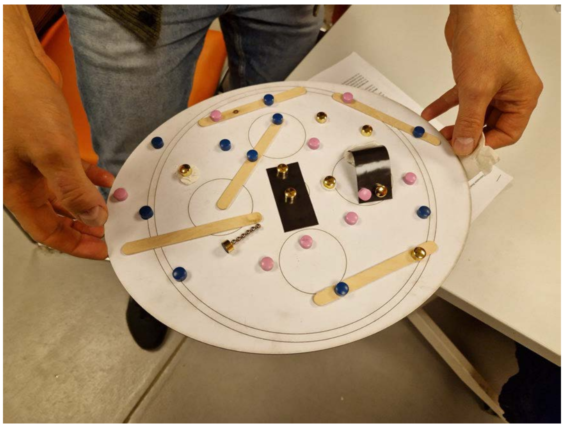


Figure 5. P1's Score.

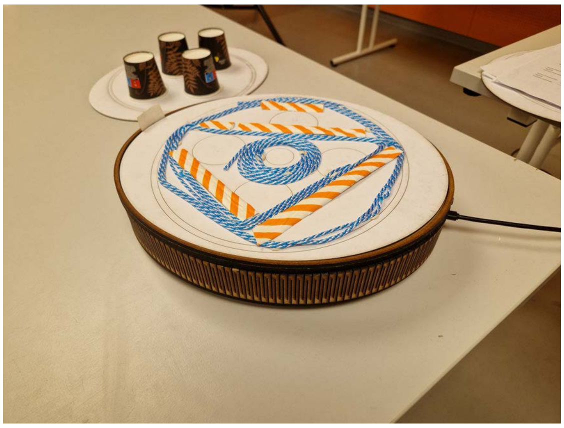


Figure 6. P2's Score.

leged mundane materials and everyday objects to free participants and facilitate the emergence of subjective aesthetic views. The exploration of these relatable and familiar objects in novel creative technology-based projects empowers and asserts confidence in new and non-specialist audiences [28].

The workshop lasted three hours. During the first hour, we briefly demoed Stacco and invited each participant to try it out using the available set of magnetic spheres. The participants were free to play with and mix all the models plus the FM synthesizer. Each of them experimented with the instrument for about ten minutes, with P2 and P4 playing more than one session.

After the break, we invited the participants to "build" their scores. We avoided using the term "compose" for its strong semantic connotation to established traditions. This choice is in line with Andersen's methodology, which avoids overly loaded terminologies to prevent participants from limiting themselves to pre-existing technological – in our case techno-musical – assumptions.

The time for the task was 50 minutes. This constraint is based on the guidance provided by Lepri and McPherson which found that a relatively fast-paced exercise helps participants sidestepping insecurity and avoid overthinking. During the remaining hour, the participants presented their works and performed by placing the ovals on top of Stacco. After the workshop, we submitted to the participants an optional online survey. The survey was completed by P1, P2 and P4.

The next paragraphs describe the strategies devised by the participants for notating their ideas and elucidate the rationales they put forward in support of their approaches.

4.2 Scores

P1 designed a one-shot composition by distributing small coloured magnets on the cardboard (Fig. 5).

The magnets were carefully distanced to not collapse into each other through their reciprocal attraction. Some of the magnets were placed on top of wooden sticks, others on top of magnetic stripes. As soon as the oval was placed on top of Stacco, its four attractors broke the balance and caused the system to collapse.

The performance consisted of this initial chaotic event and an improvisation starting from the newly found balance, aiming to elicit new unpredictable interactions.

P1 chose not to title the score. The choice of sounds, FM synthesis and magnets model, was very consequential as the composition combined the materiality of magnets with the non-linear behaviour that is typical of FM.

P2 used three-dimensional objects such as rope and straws to define separate areas of the oval through meticulous chromatic patterning (Fig. 6). The score was titled *Techno and Textiles*. The rope and the straws were chosen because of their alternating coloured and white diagonal lines. To reinforce the chromatic effect, P2 cut the straws into two-dimensional rectangles, obtaining an orange-and-white pattern, and juxtaposed multiple rope windings to create a chromatic continuity between the white and blue diagonals. In line with P1, P2 used FM and magnets' sound.

As P2 explained, the title of the score is a reflection on the relationship between techno music and textiles, two themes that are rooted and in constant dialogue in Łodz's culture, as the venues hosting techno events are for the most part repurposed textile industries.

P3 designed instead a two-dimensional graphic score using coloured markers (Fig. 7), a map sketching trajectories for four spherical magnets around Stacco's four attractors. The patterns were very diverse and altogether gave the impression of electronic schematics. The choice of the sounds fell on the choir model. As opposed to the previous participants, P3 described the FM synth sounds as "too aggressive."

P4 and P5, the ones in the group with no prior musical training, designed two three-dimensional scores (Fig. 8, 9). As P4 disclosed in the following survey, Stacco reminded him of Italian architecture, of Carlo Scarpa in particular and Nono's work "A Carlo Scarpa, architetto, ai suoi infiniti possibili" (1984). Indeed, P4's score was similar to the architectural model of a city's square, with a main structural unit represented by a cup with a stick and a rubber band holding a magnetic pendulum. On the sides, wooden sticks functioned as rails on top of the attractors for two spherical magnets. The selected sounds were choir and FM, and the performance consisted of suc-

Figure 7. P3's Score.

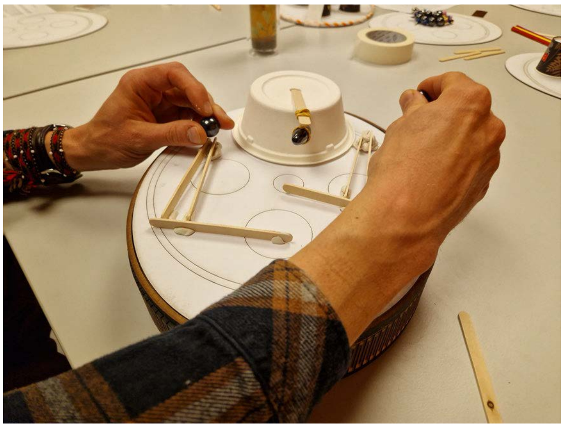


Figure 8. P4's Score.

cessive reconfigurations of the architecture, moving the spheres along the paths of the wooden rails placed on top of the attractors. Finally, P5 developed a narrative around a magnetic persona, with a larger magnetic sphere as the body and head, and arms and legs made of smaller magnets. P5 built a series of structures with straws, clay and rubber bands around three of the attractors, and titled the score *A Day in the Park*. By wandering through the score, climbing and sliding on top of straws and sticks of different heights, the character modulated Stacco's magnetic fields. As P5 noted, the work was "more as a sculpture than a practical thing." In P5's open-world score, all interactions, and in particular those elicited by gravity, are redefined as magnetism. Coherently, the sole model chosen for the performance was the magnets one.

4.3 Performances

Once the scores were completed, the participants presented them and played for the group. In this section, we will observe how the scores redefined the gestural and sonic interactions with Stacco by comparing the initial open explorations with the final performance, and including insights from the conversation that followed. This section includes observations from P1 to P4, P5 chose to explore the score alone and during the group's break. For this reason, we chose to not record this performance.

Figure 9. P5's Score.

In his exploration of Stacco, P1 started with groups of two to three spheres on each attractor, using the choir model and focusing primarily on the interactions between the spheres rather than on their correlation with the sound. After a few minutes, P1 reduced the number of spheres to two, placed each one on a different attractor and began to systematically explore the sound by lingering on the gestures that had a bigger effect. Towards the end, P1 was holding the larger spheres a few centimetres above the board and rapidly switching from one attractor to the other. As P1 noted in the survey, performing was "very interesting but sometimes random." At the same time, the interface was "gesturally sensitive." During the final discussion, P1 also noted how playing Stacco is "like fighting with the instruments." This agency, he continued, is something that every instrument has, but with Stacco "you feel the physicality of the instrument, which in electronic music you often lose." Coherently with these insights, P1 used randomness as the defining feature of the piece, letting Stacco's magnetic fields autonomously reconfigure the score.

P1's performance with Stacco was a purposeful exploration of the instrument's agency. The entanglement of the latent dimensions in RAVE was mirrored through the unpredictable entanglements of the magnets, with the score conceptually and physically mediating the dialogue between the performer and the interface. The movements during the performance were not aimed at controlling specific sound features, since, as noted during the final discussion, P1 intuitively realised their entanglement. The performative gestures were aimed instead at breaking the system's homeostasis and letting the magnetic interactions affect the sounds.

P2's test with Stacco began where P1's exploration had finished: by moving magnets from one attractor to the other and looking for emerging patterns. Once P2 found a drastic change in amplitude, he would stop and start moving one magnet around the board, looking for a gesture that could control the overall amplitude. He found a few, and one in particular producing silence, thus realising that at least one of the data points was affecting, among other parameters, the overall amplitude. P2 tried out Stacco a second time, experimenting with the cardboard before composing his piece. Through this intermediate phase he discovered a new gesture: by turning as a knob a spherical magnet placed at the centre of one attractor and pushing it firmly on the cardboard, he could fine-control the sounds better than by moving the magnets on the board. He incorporated this approach into the following performance, by pushing spheres from the centre of one attractor and then rapidly moving them to another one.

The chromatic patterning in P2's score and the title of the performance found a correspondence in the rhythmic approach to the navigation of the latent space, with the magnets model combined with the FM synthesizer to generate noisy textures.

P2 purposefully built the score around the magnetic attractors, emphasizing through colours and textures the areas of finer control. As opposed to P1, P2 used Stacco's magnetic fields as a guide in exploring the instrument-score, and represented the close relationship of textiles and techno music in the area by rhythmically displacing a single sphere from orange to blue areas and vice-versa.

P3 used instead the organ model for his performance with Stacco. He placed one sphere on each attractor and moved them around with both hands, at first slightly, then more decisively, in search of areas on the board where the magnets' movement had the most effect on the sound. The score was then sketched around these areas, with the spheres rolling back and forth from the centre of the attractors, first one by one, then together. The closure of the performance was offered by Stacco itself: P3 was aiming to place the fourth sphere on the last free circle, but the magnet got attracted by a sphere in the nearby one and slipped from P3's hand. That particular configuration in the latent space, resulting from a sudden interaction with the magnetic inscriptions, produced complete silence and an unexpected, enjoyable finale to which the whole group reacted with a laugh.

If P3's approach to the writing was rigorous and methodical, and certainly more akin than the others to the traditional idea of a graphic score than in all other scores in the session, the performance was very playful and engaging: as P3 noted in the survey, playing with Stacco related to an "attitude to experiments" that "goes back to childhood."

P4 chose to experiment with the FM synthesizer and with the choir model. While playing, P4 compared the experience of playing with the FM patch to that of a no-input mixer, emphasizing the high responsiveness to small gestures. This is a distinctive feature of FM synthesis compared to RAVE: if both algorithms are similar in the high degree of entanglement between the parameters, RAVE seems to distribute them evenly within the latent representation. The interactions between the parameters of the FM synthesizer are instead less linear, with small transitions causing marked changes. In addition to this, the higher latency of RAVE contributes to reducing the overall responsiveness to subtle and rapid gestures.

P4 explored Stacco with two magnets, moving them rapidly above and between the attractors. The gestures afforded by the score were instead very different, with spheres moving on predetermined paths. The playful character of this score emerged clearly in the way P4 manipulated the main structure and the embedded pendulum to interact with the spheres, much like a kid having fun with construction toys.

5. DISCUSSION

In this section, we describe and reflect upon themes that emerged from the workshop and the following survey, focusing in particular on how the notion of the instrument-score intersects to neural audio synthesis through embodied sketching.

5.1 Playfulness and Agency

During the demo, the participants were evidently surprised by the magnets spinning around the attractors, bumping into each other and combining into complex structures. After this initial surprise, during the practice sessions and the discussion the participants emphasised Stacco's agency as its most distinctive feature, with P1 even noticing that it felt like "fighting with the instrument," and the group laughing in witnessing Stacco's magnets calling P3's performance to an unexpected finale.

Notably, as the participants practised and acquired confidence with the system, the agency and unpredictability of the magnetic interactions, from a factor of surprise, became a compositional element, with the participants' scores leveraging it as a generator of randomness (P1), as an architectural component (P4), or as the main thread in constituting a narrative (P5).

The magnetic agency of the instrument-score, that P1's metaphor so effectively depicts, pairs with that of the algorithm in autonomously and independently re-modelling the dataset. To provide an effective way of accounting for, negotiating with, and incorporating these two features into the notation, we relied on embodied sketching.

5.2 Embodied Sketching

The approach of *Embodied Sketching*, that is, the practice of embedding the score onto the instrument itself has proven effective for different reasons.

- It favoured the transposition of the instrument's peculiarities into the notation. As participants practised, gaining a better understanding of the system, Stacco's features translated into the scores' semantics and subsequently, in P1, P2 and P5 in particular, into the choice of the sounds.
- It offered to the participants an open creative environment, in which they felt free to compose two-dimensional (P3) or three-dimensional scores (P4, P5), to adhere to traditional (P3) or novel compositional methodologies, often strongly related to their personal histories and backgrounds (P2, P4, P5) or with their personal views on composition (P1, P4).
- It allowed them to overlap latent exploration and gestural notation. This is of particular value in composing with neural synthesis, since, as we discussed in 3.1, this novel technology involves an embodied process of a posteriori exploration and a strong connection between gesture and notation.

Finally, during the discussion another theme has emerged: how different is composing and performing with neural synthesis than with a chaotic synthesizer based on a traditional synthesis technique? Even though this deserves further investigation, the participants' feedback and their consistent use of the FM patch provide some initial indications. Indeed, most of the scores combined the two methods, taking advantage of FM's higher sensitivity to small and rapid changes as well as of neural synthesis' potential for a smoother, organic sonic exploration. In future research, we aim to explore whether these two synthesis methods may suggest different notational approaches, and in what ways these might diverge.

6. CONCLUSIONS

Stacco is a new interface for a new synthesis method, a novel instrument-score that functions as a platform for composing with neural synthesis through embodied sketching, which we defined as the practice of embedding a notational layer on the instrument's body – i.e. removable oval cardboard sheets placed on top of Stacco's magnetic board.

As our workshop demonstrates, embodied sketching addresses some of the compositional and performative challenges posed by neural synthesis. Indeed, the tangible and playful interactions that Stacco entails allow to bridge gesture and notation and prioritise a holistic and embodied exploration of the model.

By "thinking with their hands" [23], rather than focusing on rationalising the model's intricate functioning, our participants intuitively developed original and highly subjective musical sketches. This was particularly clear, for instance, in P1's approach to the composition, leveraging Stacco's properties to elicit chaotic magnetic and sonic interactions, but also in P3's more traditional approach, searching for and drawing sonically satisfying trajectories around Stacco's attractors.

Despite the short duration of our workshop, the engagement with the instrument led to the creation of a variety of compelling sketches. In future workshops, we intend to investigate whether a more consistent and prolonged practice with our system might lead to a deeper embodied understanding of the model and to the creation of compelling musical works. To this end, we plan to expand Stacco's embodied musical sketches using transparent materials instead of cardboard sheets, so to stack multiple notational layers and create interchangeable scenes and modular compositions. Through this approach, and by experimenting with other inherent scores, we hope to contribute to the understanding of the creative practices and the novel semantics that the introduction of deep learning techniques into the musical and creative domains is bringing forth.

Acknowledgments

This research is supported by the European Research Council (ERC) as part of the Intelligent Instruments project (INTENT), under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 101001848).

7. REFERENCES

[1] A. Caillon and P. Esling, "Rave: A variational autoencoder for fast and high-quality neural audio synthesis," *arXiv preprint arXiv:2111.05011*, 2021.

[2] J. M. Avila, V. Tsaknaki, P. Karpashevich, C. Windlin, N. Valenti, K. Höök, A. McPherson, and S. Benford, "Soma design for nime," in *Proceedings of the 2020 International Conference on New Interfaces for Musical Expression (NIME'20)*, 2020.

[3] D. Holzer, H. Frisk, and A. Holzapfel, "Sounds of futures passed: Media archaeology and design fiction as a NIME methodologies," in *Proceedings of the Conference on New Interfaces for Musical Expression*, 2021, forthcoming.

[4] G. Lepri and A. McPherson, "Making up instruments: Design fiction for value discovery in communities of musical practice," in *Proceedings of the Designing Interactive System Conference*, 2019, pp. 113–126.

[5] F. Morreale, S. M. A. Bin, A. McPherson, P. Stapleton, and M. Wanderley, "A NIME of the times: Developing an outward-looking political agenda for this community," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2020, pp. 160–165. [Online]. Available: https://www.nime.org/proceedings/2020/nime2020_paper31.pdf

[6] M. Gurevich, "Discovering instruments in scores: A repertoire-driven approach to designing new interfaces for musical expression," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2017, pp. 163–168. [Online]. Available: http://www.nime.org/proceedings/2017/nime2017_paper0031.pdf

[7] J. Sterne, "Spectral objects: On the fetish character of music technologies," *Steintrager and Chow*, vol. 2019, pp. 94–109, 2019.

[8] M. Battier and N. Schnell, "Introducing composed instruments, technical and musicological implications," 2002.

[9] A. Lucier, "Origins of a form: Acoustical exploration, science and incessancy," *Leonardo Music Journal*, vol. 8, pp. 5–11, 1998. [Online]. Available: http://www.jstor.org/stable/1513391

[10] E. Tomás and M. Kaltenbrunner, "Tangible scores: Shaping the inherent instrument score," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. London, United Kingdom: Goldsmiths, University of London, Jun. 2014, pp. 609–614. [Online]. Available: http://www.nime.org/proceedings/2014/nime2014_352.pdf

[11] E. Tomás, *The interface-score: Electronic Musical Interface Design as Embodiment of Performance and Composition*. Linz, Kunstuniversität, 2018.

[12] N. Privato, T. Magnusson, and E. T. Einarsson, “The magnetic score: Somatosensory inscriptions and relational design in the instrument-score,” in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR’2023*, A. P. D. Ritis, V. Zappi, J. V. Buskirk, and J. Mallia, Eds. Boston, Massachusetts, USA: Northeastern University, 2023, pp. 36–44.

[13] ——, “Magnetic interactions as a somatosensory interface,” pp. 387–393, May 2023. [Online]. Available: http://nime.org/proceedings/2023/nime2023_54.pdf

[14] N. Merendino, L. G. Lepri, A. Roda, and R. Masu, “Redesigning the chowndolo: a reflection-on-action analysis to identify sustainable strategies for nimes design.” 2023.

[15] T. Cochrane, “7475On the resistance of the instrument,” in *The Emotional Power of Music: Multidisciplinary perspectives on musical arousal, expression, and social control*. Oxford University Press, 07 2013. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780199654888.003.0006

[16] K. Hook, *Designing with the body: Somaesthetic interaction design*. MIT Press, 2018.

[17] D. Rocchesso, D. A. Mauro, and S. D. Monache, “Mimic: The microphone as a pencil,” ser. TEI ’16. Association for Computing Machinery, 2016, pp. 357–364. [Online]. Available: https://doi.org/10.1145/2839462.2839467

[18] S. Delle Monache, D. Rocchesso, F. Bevilacqua, G. Lemaitre, S. Baldan, and A. Cera, “Embodied sound design,” *International Journal of Human-Computer Studies*, vol. 118, pp. 47–59, 2018.

[19] A. Camurri, G. Volpe, S. Piana, M. Mancini, R. Niewiadomski, N. Ferrari, and C. Canepa, “The dancer in the eye: towards a multi-layered computational framework of qualities in movement,” in *Proceedings of the 3rd International Symposium on Movement and Computing*, 2016, pp. 1–7.

[20] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe, “Eyesweb: Toward gesture and affect recognition in interactive dance and music systems,” *Computer Music Journal*, vol. 24, no. 1, pp. 57–69, 2000.

[21] K. Andersen and R. Wakkary, “The magic machine workshops: Making personal design knowledge,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[22] A. Boal, *Games for Actors and Non Actors, tr. A. Jackson*. London, Routledge, 1992.

[23] K. Andersen, “Making magic machines,” in *10th European Academy of Design Conference*, 2013.

[24] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” *CoRR*, vol. abs/1609.03499, 2016. [Online]. Available: http://arxiv.org/abs/1609.03499

[25] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, and Y. Bengio, “Samplernn: An unconditional end-to-end neural audio generation model,” 2017.

[26] C. Dobrian and D. Koppelman, “The ’e’ in nime: Musical expression with new computer interfaces.” 01 2006, pp. 277–282.

[27] G. Moro, S. Bin, R. Jack, C. Heinrichs, and A. Mcpherson, “Making high-performance embedded instruments with bela and pure data,” 06 2016.

[28] G. Lepri, J. Bowers, S. Topley, P. Stapleton, P. Bennett, K. Andersen, and A. McPherson, “The 10,000 instruments workshop - (im)practical research for critical speculation,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2022. [Online]. Available: https://doi.org/10.21428%2F92fbeb44.9e7c9ba3

TOKENIZATION OF MIDI SEQUENCES FOR TRANSCRIPTION

Florent Jacquemard
INRIA/CNAM
florent.jacquemard@inria.fr

Masahiko Sakai
Nagoya University
sakai@i.nagoya-u.ac.jp

Yosuke Amagasu
Nagoya University
amagasu@trs.css.i.nagoya-u.ac.jp

ABSTRACT

There generally exists no simple one-to-one relationship between the events of a MIDI sequence, such as note-on and note-off messages, and the corresponding music notation elements, such as notes, rests, chords, and ornaments.

We propose a method for building a formal correspondence between them through a notion of tokens in an input MIDI event sequence and an effective tokenization approach based on a hierarchical representation of music scores. Our tokenization procedure is integrated with an algorithm for music transcription based on parsing *wrt* a weighted tree grammar. Its effectiveness is shown in examples.

1. INTRODUCTION

In the context of compiling (lexical analysis) and in Natural Language Processing (NLP), tokenization is the act of dividing a character sequence into elementary subsequences called tokens. The primary purpose of tokenization is to identify, in the early steps of processing, meaningful subsequences of characters in preparation for the next steps. For instance, tokenization consists of finding keywords and identifiers with automata-based pattern-matching techniques in compiling. For NLP based on statistical learning, tokenization aims at building sequential encodings of character sequences into integer vectors, using a dictionary of tokens, for training language models.

In languages based on Latin alphabets (including programming languages), tokens are generally words or subwords. The tokenization into words, in particular, is eased by space separation. The case of written Japanese language is more complicated since there is no space separation (although some punctuation exists). Therefore, some real-time tokenization process is required when reading a text. In the case of hiragana characters (phonetic characters), ambiguity may arise in tokenization, leading to confusion.

Similar ambiguity problems may occur when processing symbolic music data in sequential form, in particular MIDI data, in order to extract more structured information. A MIDI flow (or file) is essentially an unstructured sequence of *messages*, each one representing an elementary *event* for the production of music with an electronic instrument. Two common events are the start of a note (*note-on*), corresponding to a key press, and the end of note (*note-off*), for a key release. On the other hand, higher-level musical elements of Common Western musical notation, such as notes, rests, chords, grace notes, and other ornaments may correspond to several successive MIDI events, and conversely. Therefore, extracting such notational elements from a MIDI flow, in *e.g.*, a task like MIDI-to-score transcription, requires a processing similar to tokenization.

Let us illustrate this point of view with sneak observations on the short MIDI sequence of Figure 1, and some of its possible denotations as a 4/4 measure in a music score. We use a representation similar to a piano-roll, with note pitches on the vertical axis, time on the horizontal axis, and note-on events depicted as black dots and note-off as white dots. Several grouping of these events into *tokens* are considered in cases (a) to (h) of Figure 1 and correspond to different notations. Note that the piano-roll is the same in all these cases. In each case, the tokens, called T_0 to T_n (n depending on the case) are represented with blue boxes. Intuitively, all events in a token T_i represent one or several music notation elements occurring at the same time position τ_i in a score. The time position τ_i associated with the token T_i is indicated on the axis under the corresponding blue box. The notation elements represented may be *e.g.*, one note, one chord, some grace notes and one note, *etc*. In other terms, the groupings defined by the tokens can be seen as a re-alignment of MIDI events, like with a rhythm quantization functionality of a Digital Audio Workstation (DAW), except that the alignment points τ_i are not evenly distributed in a grid. One strategy for defining the alignment points τ_i according to hierarchical structures is proposed in Section 4.

Going back to Figure 1, in the first tokenization (a), and corresponding notation (a)', the two first note-on events (D4 and A4) belong both to the first token T_0, and their matching note-off events belong to the next tokens, resp. T_1 and T_2. This situation corresponds to two notes starting simultaneously on the first beat of the measure (time position $\tau_0 = 0$), and ending on different later beats: beat 2 (time position $\tau_1 = \frac{1}{4}$) for D4, and beat 3 (time position $\tau_2 = \frac{1}{2}$) for A4. It is denoted as an interval tied to a quarter note.

In the tokenization in Figure1(f) however, both the note-on and note-off events of D4 belong to the first token T_0, aligned at time position $\tau_0 = 0$ (beat 1). It follows that this note is a grace-note located at τ_0. On the opposite, the A4 has its note-on in T_0 and its note-off in T_1, aligned at time

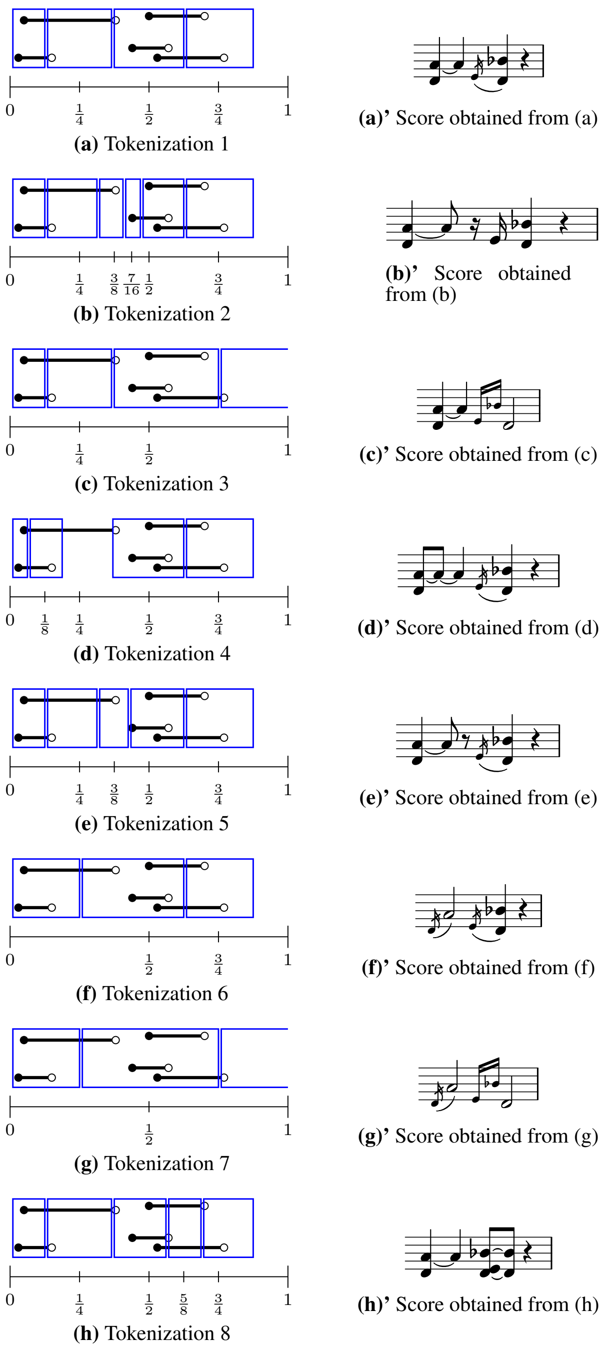


Figure 1. Examples of MIDI tokenizations and transcriptions.

position $\tau_1 = \frac{1}{2}$ (beat 3). Hence, this note is a half note, with the D4 as an ornament, see Figure 1(f)'.

The interpretation of the other tokens of Fig. 1 is similar. Note that in case (c), and (g), two pairs of matching note-on and note-off events are embedded in the third token. It follows that the last note D4 has two grace notes E and B♭.

Therefore, we see in Figure 1 many possible denotations of the same MIDI sequence. How were chosen the tokens in this picture, in order to be associated a notation? Which denotation shall be considered the best?

In this paper, we propose a formal definition of the notion of token of MIDI events, as an attempt to clarify the correspondence between MIDI events and music notation elements. We characterize in particular what notation can be associated with a token, defining several *types* of tokens, corresponding to different notations. We focus in particular on notation elements which are not the most studied in the literature on music notation and symbolic music processing in general, namely the *rests* and the *ornaments*.

The main motivation for this study is the processing of MIDI data, in particular its transcription into Common Western music notation, a task sometimes referred to as MIDI-to-score Automatic Music Transcription (M2S-AMT). We propose an efficient procedure for tokenization and its integration into a M2S-AMT framework. The approach is based on *parsing* [1], in its usual computational sense: the inferrence of some (hierarchical) structure from unstructured (sequential) data. In these settings, we consider tree structures for defining hierarchically the time boundaries of tokens. A weighted tree grammar generates those trees, and the estimation of the best tree (*wrt* weight) in the grammar's language is performed by a dynamic programming algorithm. This method is implemented in a framework for MIDI-to-score transcription in development.

Related Work. Compiling (textual) programming languages generally involves steps of lexical analysis and syntactic analysis (parsing) [1]. In some sense, we follow a similar approach in our transcription approach, briefly described above. However, a significant difference is that, in our case, several tokenizations of the MIDI input are possible, and we must explore possibilities to choose the one giving the best notational result.

In the literature on music generation, several works propose sequential encodings of MIDI data in order to train models like the Transformer [2, 3, 4, 5, 6, 7, 8, 9, 10]. Some of these encodings, often referred to as *MIDI tokenization*, are implemented in the library MIDItok [11]. Their purpose is to format MIDI data as a sequence of tokens, where, roughly, each token represents one elementary component of a MIDI event (position, duration, pitch value, *etc*). This notion of elementary token is very different from the one we present here, where every token is made of several MIDI events (considered atomic) and conveys higher-level musical information about the notation. Some strategies for grouping elementary tokens are proposed *e.g.*, in [7], in order to reduce training sequences. Another critical difference is that the purpose of the above tokenization approaches is the generation of MIDI data by trained model, whereas our purpose is the construction of music notation from MIDI data.

The sequential input and structured output of M2S-AMT are of very different natures, and establishing a relationship between input MIDI events and output score elements is not a trivial task, considering the richness of music notation. For instance, many existing M2S transcription tools, including general public ones, like MuseScore, fall short of detecting ornaments and rests. As observed formarly [12] problems often come from mismatches between input and output. Several transcription tools ignore note-off events to cope with the mismatches. In other works, note-on and note-off events are treated in different passes [13].

We believe that the notions of *tokens* and *token types* proposed here can help in making progress in the production of music notation, such as in M2S-AMT, in particular regarding the treatment of rests and ornaments in preprocessing. However, it should be noted that in the present work, we only consider homophonic input (monophonic voices including chords), whereas [13, 14] deal with the transcription piano input, which is a much harder problem.

The main contributions of this paper are the following:

- the definition of *tokens* as sequences of MIDI events corresponding to notation elements (Section 2),
- the proposition of an efficient procedure using trees for dividing a MIDI sequence into tokens (Section 3),
- a dynamic programming algorithm for parsing a MIDI sequence into structured music notation, joined with the above tokenization procedure (Section 4).

2. MIDI EVENTS, TOKENS AND SCORE ELEMENTS

In this section, we define a correspondence between music notation elements and sequences of MIDI events called *tokens*, as described informally in Figure 1. Each token is given a token type, a note, rest, or chord, with or without an ornament, corresponding to the score element. On the other hand, a MIDI event in a token may have an individual role, which we call an *event role in the token*.

2.1 MIDI Events

We consider a typical representation, in the form of so-called timestamped MIDI events, of the note-on and note-off messages in a MIDI file, often depicted as a piano-roll. A MIDI event e (or *event* for short) is made of the following components:

- a *real-time value* $\mathsf{ts}(e) \in \mathbb{Q}$, expressed in seconds,
- an integral *pitch value* $\mathsf{pitch}(e) \in \{0, \ldots, 128\}$,
- an integral *velocity value* $\mathsf{vel}(e) \in \{0, \ldots, 128\}$,
- an attribute $\mathsf{flag}(e)$ which is either on when e is a *note-on* message, or off for a *note-off* message.

A MIDI *sequence* $E = \langle e_1, e_2, \ldots, e_n \rangle$ is a finite sequence of MIDI events with increasing timestamps, *i.e.*, such that $\mathsf{ts}(e_1) \leq \mathsf{ts}(e_2) \leq \cdots \leq \mathsf{ts}(e_n)$. The length of E is denoted by $|E| = n$ and the *concatenation* of two sequences E and E' is denoted by $E\,E'$. We use the set-like-notation $e \in E$ to express that the event e belongs to the MIDI sequence E. The subset of note-on (resp. note-off) events in a sequence E of events is denoted by $\mathsf{on}(E) = \{e \in E \mid \mathsf{flag}(e) = \mathsf{on}\}$ (resp. $\mathsf{off}(E) = \{e \in E \mid \mathsf{flag}(e) = \mathsf{off}\}$).

The event *matching* e in a sequence E, denoted by e^{-1} is the first note-off after e with the same pitch when e is a note-on, and the note-on event d such that $d^{-1} = e$ when e is a note-off. Formally, for all event e,

- $\mathsf{pitch}(e^{-1}) = \mathsf{pitch}(e)$ and $\mathsf{flag}(e^{-1}) \neq \mathsf{flag}(e)$,
- if $\mathsf{flag}(e) = \mathsf{on}$, then $\mathsf{ts}(e) < \mathsf{ts}(e^{-1})$ and for all e' such that $\mathsf{ts}(e) < \mathsf{ts}(e') < \mathsf{ts}(e^{-1})$, it holds that $\mathsf{pitch}(e') \neq \mathsf{pitch}(e)$, and
- if $\mathsf{flag}(e) = \mathsf{off}$, then $e^{-1} = d$ such that $d^{-1} = e$.

The matching operator is idempotent: $(e^{-1})^{-1} = e$. We shall consider below only well-formed MIDI sequences E, such that e^{-1} exists in E for all $e \in E$.

Example 1. *Figure 2 presents the events $e_1, \ldots, e_{10}$ of the MIDI sequence in Figure 1. This sequence is well-formed, and ${e_1}^{-1} = e_3$, ${e_2}^{-1} = e_4$, ${e_3}^{-1} = e_1$, ${e_4}^{-1} = e_2$, ${e_5}^{-1} = e_8, \ldots, {e_{10}}^{-1} = e_7$.*

2.2 Tokens

In order to be converted into music notation, the MIDI events of an input sequence can be aligned to some salient time points, like the *grid points* in DAWs or in earlier rhythm quantization algorithms [15, 16]. It may happen that several events, neighbours in a MIDI sequence E are aligned to the same time point. The notion of token aims at capturing the subsequences of events of E made simultaneous after alignment. More precisely, a *token* of E is a non-empty subsequence $T = \langle e_i, \ldots, e_j \rangle$ of E with $1 \leq i \leq j \leq n$, that is moreover closed *wrt* time equality: every $e \in E$ such that $\mathsf{ts}(e) = \mathsf{ts}(e_k)$ for some $i \leq k \leq j$ must be in T.

Intuitively, a token aims at containing some MIDI events representing simultaneous score elements, *i.e.*, elements occurring at the same musical date (beat) in the score. This can be the case of several notes involved in a chord, or of one note with one or several grace notes, or another specific ornament (*mordent*, *gruppetto*, *trill*...). Note that we are talking here about the theoretical simultaneity of notes in the score, not of simultaneity in a performance. In the score, all the notes of an ornament occur at the same theoretical date, in beats, as the note they decorate. However, during a performance, they will occur shortly before or after the decorated note.

2.3 Event Role in a Token

Only some combinations of MIDI events will appropriately fit into one case, corresponding to a notation, like in Figure 1. Having too many events in a token may make no sense. In order to capture tokens that can be considered valid, *i.e.*, that can be transcribed into simultaneous notation elements, we introduce the notion of *role* of a MIDI event e in a token T. It is defined according to the flag of e (on or off-note) and to the presence or not in T of the matching event e^{-1}, as presented in Figure 3.

Intuitively, the role *grace note* corresponds to notes with a theoretical duration of 0 in a score. It can be an *appogiatura*, an *accaciatura*, or a part of another ornament, which expresses the fact that both the start (note-on e) and the end (note-off e^{-1}) of the corresponding note belong to the same token T.

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
ts	0.03	0.05	0.15	0.38	0.44	0.50	0.53	0.57	0.70	0.77
flag	on	on	off	off	on	on	on	off	off	off
pitch	D4	A4	D4	A4	E4	B♭4	D4	E4	B♭4	D4
vel	110	100	0	0	45	90	110	0	0	0

Here, MIDI pitches are presented by note name and octave; D4 = 62, E4 = 64, A4 = 69, and B♭4 = 70.

Figure 2. MIDI event sequence for Figure 1.

	$e^{-1} \in T$	$e^{-1} \notin T$
$e \in \mathsf{on}(T)$	*grace-note*	*note*
$e \in \mathsf{off}(T)$	*goff*	*noff*

Figure 3. Role of event e in token T.

On the opposite, an event of role *note* corresponds to a note starting in a token T ($e \in \mathsf{on}(T)$), but ending in a subsequent token ($e^{-1} \notin T$).

A note-off event matching an event in the token will have role *goff*. Otherwise, it shall have role *noff*. The event role *goff* is important in the context of MIDI processing. Indeed, events with this role correspond to *micro-rests* that may occur *e.g.*, when a key of a MIDI keyboard is released before the key for the next note is pressed (*i.e.*, the play is not enough *legato*). In general, one does not want such artifacts to be transcribed literally, and in several transcription procedures, micro-rests are discarded in a pre-processing step relying on parameters to be set, like the maximal duration of a micro-rest. In our approach, micro-rests are detected as a special case of event role *noff* or *goff* in a token, and may be discarded at will.

Example 2. *The roles of events in the tokens represented in (a) and (b) of Figure 1 are displayed in Figure 4.*

2.4 Token Types and Validity

We shall use the number $ns_E(T)$ of notes sounding just after the timestamp ℓ of the last event of a token T. Formally, it is defined as the following cardinality:

$$ns_E(T) = \left|\left\{e \in E \mid \mathsf{flag}(e) = \mathsf{on}, \mathsf{ts}(e) \leq \ell < \mathsf{ts}(e^{-1})\right\}\right|$$

where $\ell = \mathsf{ts}(e)$ for the last event $e \in T$ *i.e.*, for all $e' \in T$, $\mathsf{ts}(e') \leq \mathsf{ts}(e)$.

Example 3. *The number of notes sounding just after the tokens represented in (a) and (b) of Figure 1 are displayed in Figure 5.*

We consider a finite set $\mathcal{K}$ of type names that can be assigned a token T in a MIDI sequence E as follows, according to the respective roles of the events in T:

T has type *chord* with n (> 0) notes and an *ornament* of size p (≥ 0), denoted by $\mathsf{ch}_{n,p}$, if and only if

- it contains n events of role *note*,
- it contains p events of role *grace-note*,
- each *grace note* occurs before any *note* in T, and
- $ns_E(T) = n$ (the n notes sound after the token).

T has type *rest* denoted by r, if and only if $T = \mathsf{off}(T) (\neq \emptyset)$ and $ns_E(T) = 0$,

T has type *partial continuation*, denoted by pc, if and only if $T = \mathsf{off}(T) (\neq \emptyset)$, and $ns_E(T) > 0$.

The cases of a single note or an interval correspond respectively to the types *chord* with 1 note and *chord* with 2 notes. We use the name *chord* for these types of abuse in order to shorten the definition of token types. As explained above, an ornament of size p corresponds to a sequence of events occurring at the same musical date (beat) in a score. In this description, we use, for simplicity, the generic term of "*ornament of size* p", but we could distinguish between different kinds of ornaments, such as *mordent*, *gruppetto* or *trill*, or also a *tremolo* (or *roll* in the case of drums), by analyzing the respective MIDI pitch values of the events involved.

A *rest* corresponds to the case where some note ends during the token T (a note-off $e \in T$). No other note or grace note shall start in the same token (the condition $T = \mathsf{off}(T)$ implies that $\mathsf{on}(T) = \emptyset$), and every note started before ends in T (condition $ns_E(T) = 0$). Moreover, T contains no grace note (another consequence of $\mathsf{on}(T) = \emptyset$), since in CW music notation, no grace notes or ornaments may be attached to rests. The latter case may, however, correspond to a note played very shortly, and wrongly interpreted as a grace note. In our framework, we consider an option where such a grace note is transcribed into a note with an articulation *staccato* or *staccatissimo*.

A *continuation* represents the prolongation of an event (either a note or a chord), with a tie or a dot. A continuation is *partial* when some but not all chord notes are prolongated. Remark that no token exists with type (full) continuation because we defined an empty set, which corresponds to a continuation, is not a token.

2.5 Token Validity

Tokens of the above types will be considered *valid* or not for transcription, according to the case of input considered.

2.5.1 Monophonic case

When the input is considered as strictly monophonic, at most one note shall sound at a time. Hence, a token considered valid in this case is a note, *i.e.*, a chord with 1 note, with an ornament of size $p \geq 0$ or a rest. This case is appropriate *e.g.*, for the M2S transcription of a single voice, of a monophonic instrument like winds or one piano voice in strict counterpoint.

	Figure 1(a)										Figure 1(b)									
event	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
role	n	n	no	no	gn	n	n	go	no	no	n	n	no	no	n	n	n	no	no	no
token	T_0	T_0	T_1	T_2	T_2	T_2	T_2	T_2	T_3	T_3	T_0	T_0	T_1	T_2	T_3	T_4	T_4	T_4	T_5	T_5

where $T_0, T_1, \ldots$ are the displayed tokens from left to right in the corresponding figures.

Figure 4. Event roles in some tokens of Figure 1.

	Figure 1(a)				Figure 1(b)					
token	T_0	T_1	T_2	T_3	T_0	T_1	T_2	T_3	T_4	T_5
$ns_E(T_i)$	2	1	2	0	2	1	0	1	2	0

Figure 5. Values of ns_E for some tokens of Figure 1.

2.5.2 *Homophonic case*

This case generalizes the previous one by allowing some chords in a monophonic voice. Then, tokens of all three above types are considered valid, and the others are not valid (discarded in the transcription procedure).
This case is appropriate for the M2S transcription of monophonic instruments like strings (where some chords may occur) or one voice of the piano.

2.5.3 *Case of Drums*

In the case of MIDI files recorded with electronic drumkits, see *e.g.*, [17, 18], the MIDI pitches encode the drumkit parts (snare drum, kick, toms, hi-hat, and different kinds of cymbals) and the play mode (*e.g.*, on the head or rim of snare drum or toms, on the ride, edge, or bow of cymbals, and so on) Moreover, in such input, the note-off events are not significant (for instance, they are added at a fixed duration, *e.g.*, 20ms, after the matching note-on event) and can be safely ignored. Concretely, for this case, we change the definition of event role such that every $e \in \mathsf{off}(T)$ has an extra role *ignored*, which excludes the rests from the transcription output. Actually, in a post-processing step, some rests are re-introduced by conversion of some continuations, when appropriate for the readability of the score (see [18] for details).

The definition of token validity follows additional constraints for drums. First, only ornaments of size 1, called *flams*, are allowed, and only for some parts of the drumkit. Second, chords are allowed (the drum is a polyphonic instrument) but with some restrictions: sticks cannot play more than two notes, and two notes played by pedals involved in the same chord. Finally, some particular effects like *cross-stick*, *rim-shot* or *buzz-roll* (see descriptions in [18]) require a particular processing of tokens. Some possible errors of capitation by MIDI sensors also need to be handled. We leave the details about these particular cases out of the scope of this paper.

2.5.4 *Polyphonic case*

This paper does not cover the polyphonic case, which requires a step of voice separation. To directly cope with polyphonic MIDI sequences, tokens are not just a partition of the input sequence. In fact, multiple tokens shall correspond to interleaved events in a MIDI-sequence, which is a complex extension to handle.

We call *tokenizer* a procedure that takes in input a MIDI sequence E and returns a sequence of valid tokens $T_1, \ldots, T_k$ that form a partition of E.

3. TREE-BASED TOKENIZATION PROCEDURE

In this section, we propose a tokenizer based on a tree structure. Intuitively, the idea is that the salient time points defining the tokens (by alignment of MIDI events) are characterized by a recursive subdivision of time intervals to reflect a metric organization of musical events [19]. This point of view is more involved (and more realistic) than considering regular space points in a grid (*e.g.*, one point every 16th note). It corresponds to a tree-based representation of music notation.

3.1 Real and Musical Time Scales

Here, we consider input MIDI sequences corresponding to performances. Their MIDI events are timestamped in a *real time* scale, whose units are seconds. By contrast, the durations in a music score are expressed in a *musical time* scale, whose units are beats or bars (given a time signature). The correspondence between these two time-scales is ensured by a tempo function, converting real-time values to musical-time ones. For the sake of presentation, we assume here a coincidence of these two-time scales, or equivalently, a constant tempo. We especially use a bar as a unit of musical time. The extension of the approach presented here to varying tempo functions is out of the scope of this paper.

3.2 Time Intervals

A *time interval* $I = [\tau_0, \tau_1)$ is defined by a pair of time points $\tau_0 \in \mathbb{Q}$ and $\tau_1 \in \mathbb{Q} \cup \{+\infty\}$: I starts at time τ_0 and ends just before τ_1, *i.e.*, $[\tau_0, \tau_1) = \{\tau \mid \tau_0 \leq \tau < \tau_1\}$. In the sequel, we use the notation $start(I)$ for τ_0 and $end(I)$ for τ_1. The interval $div_n(I, i)$, for $1 \leq i \leq n$, is the ith sub-interval obtained by splitting I into n parts of equal duration $\Delta = \dfrac{\tau_1 - \tau_0}{n}$ when $I = [\tau_0, \tau_1)$,

$$div_n(I, i) = [\tau_0 + (i-1) \cdot \Delta,\ \tau_0 + i \cdot \Delta).$$

If $end(I) = +\infty$, then $div_n(I, i)$ is undefined. Moreover, for $I = [\tau_0, \tau_1)$ with $\tau_1 > \tau_0 + 1$ (including the case $\tau_1 = +\infty$), we define $unit(I)$ as the subinterval $[\tau_0, \tau_0+1)$ of duration one bar, and $rem(I)$ for the remaining interval $[\tau_0 + 1, \tau_1)$, *i.e.*, $rem(I) = I \setminus unit(I)$. The latter is undefined when $\tau_1 < \tau_0 + 1$.

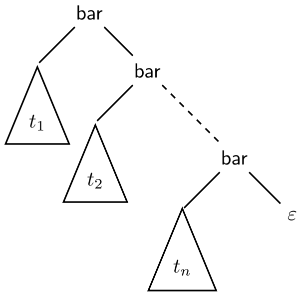


Figure 6. The form of trees

3.3 Trees

The trees defined in this section represent hierarchies of nested time intervals, defining salient points for alignment in tokens. Intuitively, the higher the points are in the hierarchy (in the tree), the strongest the corresponding beat is, from a meter point of view. The trees are labeled with two kinds of symbols:

- every inner node is labeled with a function symbol from a fixed finite set $\mathcal{F}$,
- every leaf is a node labeled with the empty symbol ε or a token type in $\mathcal{K}$, as defined in Section 2.4.

Every function symbol of $\mathcal{F}$ is associated with an operation on time intervals. We consider here two kinds of such operations:

div_n, for $n > 0$, divides an interval I into n sub-intervals $div_n(I, 1), \ldots, div_n(I, n)$ of the same duration;

bar (*bar split*), divides an interval I in two parts: $unit(I)$ and $rem(I)$.

Here, trees are in the form of Figure 6. Each left child of bar represents the content of one measure. We restrict the second argument of the bottom bar to ε for termination.

The set of trees labeled with symbols of $\mathcal{F}$ and $\mathcal{K}$ is denoted by $\mathcal{T}(\mathcal{F}, \mathcal{K})$. We denote by $yield(t)$ the sequence of token type names of $\mathcal{K}$ labelling the leaves of a tree $t \in \mathcal{T}(\mathcal{F}, \mathcal{K})$. Note that $yield(t)$ does not contain an empty leaf symbol ε.

Example 4. *Figure 7 depicts the trees corresponding to the scores in cases (a) to (d) of Figure 1. Each tree, rooted with* bar*, represents one measure whose content is the left subtree.*

3.4 Tree-based Tokenizer

Let $t \in \mathcal{T}(\mathcal{F}, \mathcal{K})$ and let I be a time interval associated with the root node of t. We can associate a sub-interval of I to every other node of t, following the above interpretation of the symbols of $\mathcal{F}$.

Example 5. *In Figure 7, the time intervals associated with each node of trees are represented in red. They are computed from the interval assigned to the root node by recursive application of the operators labeling the nodes.*

Formally, the interval assignment is defined as follows, where $itv(t, I)$ is the sequence of the intervals labelling the leaves of t.

$$\begin{aligned}
itv\big(\mathsf{div}_n(t_1, \ldots, t_n), I\big) &= itv(t_1, div_n(I, 1)) \ldots \\
&\quad itv(t_n, div_n(I, n)) \qquad \text{if } n > 0 \text{ and } end(I) \in \mathbb{Q}, \\
itv\big(\mathsf{bar}(t_1, t_2), I\big) &= itv(t_1, unit(I))\, itv(t_2, rem(I)) \\
&\qquad\qquad \text{if } end(I) = +\infty, \\
itv\big(\theta, I\big) &= I \qquad \text{for } \theta \in \mathcal{K} \cup \{\varepsilon\}.
\end{aligned}$$

It can be observed that, when it is defined, $itv(t, I)$ forms a partition of I. It is always defined when $I = [\tau_0, +\infty)$ and t has the form of a right combination.
The sequence of time points induced by the partition $itv(t, I)$ is called the ***grid*** of t of carrier I, denoted by $grid(t, I)$. Formally, if $itv(t, I)$ is a partition of the form $[\tau_0, \tau_1), [\tau_1, \tau_2), \ldots, [\tau_{k-1}, \tau_k)$ with $\tau_0 = start(I)$ and $\tau_k = end(I)$, for some $k \geq 0$ (τ_k might be $+\infty$), then:

$$grid(t, I) = \langle \tau_0, \ldots, \tau_k \rangle$$

To every time point τ_i in the grid, we can associate a set T_i of events in E containing the MIDI events closer to τ_i than to its neighbours τ_{i-1} and τ_{i+1} in the grid. For a formal definition, we introduce a notation $cp(\tau_1, \tau_2)$ to the present center point of τ_1 and τ_2:

$$cp(\tau_1, \tau_2) = \tau_1 + \frac{\tau_2 - \tau_1}{2} = \frac{\tau_1 + \tau_2}{2}$$

$$\begin{aligned}
T_0 &= \big\{ e \in E \mid \tau_0 \leq \mathsf{ts}(e) < cp(\tau_0, \tau_1) \big\}, \\
T_i &= \big\{ e \in E \mid\ cp(\tau_{i-1}, \tau_i) \leq \mathsf{ts}(e) < cp(\tau_i, \tau_{i+1})\ \big\} \\
&\qquad\qquad \text{for all } i, 0 < i < k.
\end{aligned}$$

Based on $grid(t, I)$ and the above sequence $\langle T_0, \ldots, T_{k-1} \rangle$, we define

$$tokenize(t, I) = \langle (T_{i_0}, \tau_{i_0}), \ldots, (T_{i_p}, \tau_{i_p}) \rangle$$

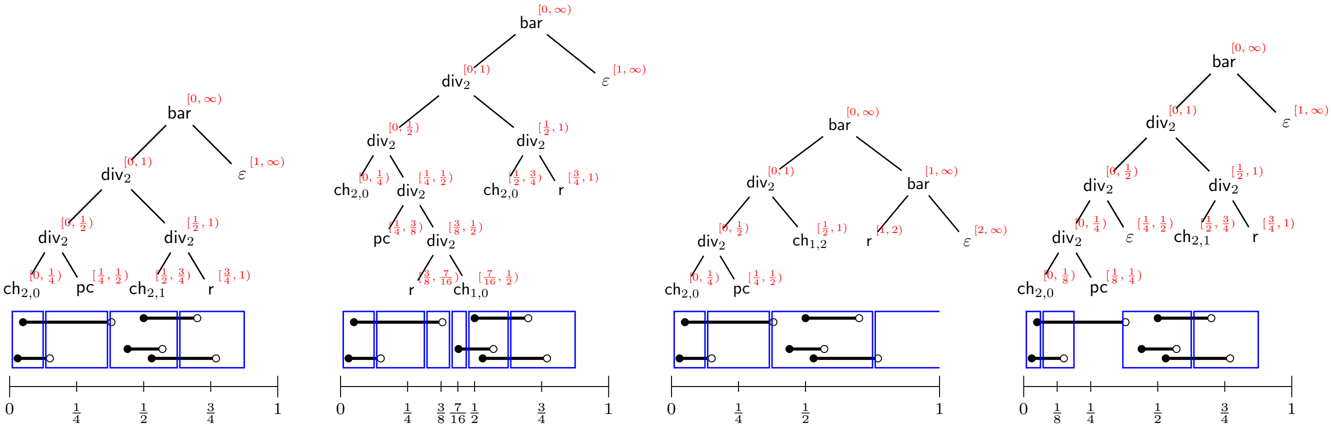
where $0 \leq i_0 < \ldots < i_p \leq k - 1$ is the sequence of indices $0 \leq j \leq p$ such that $T_{i_j} \neq \emptyset$. In the sequel, we use a renumbered tokenized sequence as

$$tokenize(t, I) = \langle (T_0, \tau_0), \ldots, (T_p, \tau_p) \rangle.$$

Example 6. *For the trees of Figure 7 it holds that:*

$$\begin{aligned}
grid\big(t_1, [0, \infty)\big) &= \langle 0, \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{3}{4}, 1, \infty \rangle, \\
tokenize\big(t_1, [0, \infty)\big) &= \langle (T_0, 0), (T_1, \tfrac{1}{4}), (T_2, \tfrac{1}{2}), (T_3, \tfrac{3}{4}) \rangle, \\
grid\big(t_2, [0, \infty)\big) &= \langle 0, \tfrac{1}{4}, \tfrac{3}{8}, \tfrac{7}{16}, \tfrac{1}{2}, \tfrac{3}{4}, 1, \infty \rangle, \\
tokenize\big(t_2, [0, \infty)\big) &= \\
&\langle (T_0, 0), (T_1, \tfrac{1}{4}), (T_2, \tfrac{3}{8}), (T_3, \tfrac{7}{16}), (T_4, \tfrac{1}{2}), (T_5, \tfrac{3}{4}) \rangle, \\
grid\big(t_3, [0, \infty)\big) &= \langle 0, \tfrac{1}{4}, \tfrac{1}{2}, 1, 2, +\infty \rangle \\
tokenize\big(t_3, [0, \infty)\big) &= \langle (T_0, 0), (T_1, \tfrac{1}{4}), (T_2, \tfrac{1}{2}), (T_3, 1) \rangle \\
grid\big(t_4, [0, \infty)\big) &= \langle 0, \tfrac{1}{8}, \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{3}{4}, 1, +\infty \rangle, \\
tokenize\big(t_4, [0, \infty)\big) &= \langle (T_0, 0), (T_1, \tfrac{1}{8}), (T_2, \tfrac{1}{2}), (T_3, \tfrac{3}{4}) \rangle,
\end{aligned}$$

where $T_0, T_1, \ldots$ are the displayed tokens from left to right in the corresponding figures.



(a) Tree t_1 for Figure 1(a) **(b)** Tree t_2 for Figure 1(b) **(c)** Tree t_3 for Figure 1(c) **(d)** Tree t_4 for Figure 1(d)

Figure 7. Rhythm Trees for Figure 1

4. TREE LANGUAGES, PARSING AND M2S-TRANSCRIPTION

In this section, we propose to use weighted tree grammar in order to define languages of trees of Section 3.3. Then we reduce M2S-AMT into the problem of parsing an input MIDI sequence *wrt* such a grammar.

4.1 Tree Grammar and Cost of Readability

A weighted tree grammar is a tuple $\mathcal{G} = \langle \mathcal{N}, A_0, \mathcal{F}, \mathcal{K}, \mathcal{R} \rangle$ where $\mathcal{N}$ is a finite set of non-terminal symbols, $A_0 \in \mathcal{N}$ is the initial non-terminal, $\mathcal{F}$ and $\mathcal{K}$ are as in Section 3.3, and $\mathcal{R}$ is a finite set of *weighted production rules*. Every production rule of $\mathcal{R}$ has one of the following forms:

$A \xrightarrow{w} f(A_1, \ldots, A_n)$
where $A, A_1, \ldots, A_n \in \mathcal{N}$, $f \in \mathcal{F}$, and $w \in \mathbb{Q}$,
$A \xrightarrow{w} \theta \mid \varepsilon$
where $A \in \mathcal{N}$, $\theta \in \mathcal{K}$ and $w \in \mathbb{Q}$.

For each rule ρ of one of the two above kinds, $w \in \mathbb{Q}$ is called the weight of ρ, denoted by $weight(\rho) = w$. It is used to compute a value of tree complexity (called *cost of readability*). With the rules of the second kind, a readability cost value w is associated with every $\theta \in \mathcal{K}$ and non terminal A.
Let us extend $\mathcal{T}(\mathcal{F}, \mathcal{K} \cup \{\varepsilon\})$ to $\mathcal{T}(\mathcal{F}, \mathcal{K} \cup \{\varepsilon\} \cup \mathcal{N})$, the set of trees whose leaves can be labeled with type names of $\mathcal{K}$ or non-terminals of $\mathcal{N}$. A derivation D of $\mathcal{G}$ is a sequence of the form: $A_0 \xrightarrow{\rho_1} t_1 \xrightarrow{\rho_2} \ldots t_{k-1} \xrightarrow{\rho_k} t_k$, where, for all $0 < i \leq k$, $t_{i-1}, t_i \in \mathcal{T}(\mathcal{F}, \mathcal{K} \cup \{\varepsilon\} \cup \mathcal{N})$, $\rho_i = A_i \rightarrow u_i \in \mathcal{R}$, the leftmost innermost occurence of a non-terminal in t_{i-1} is A_i and t_i is obtained from t_{i-1} by replacing this occurence of A_i by u_i. We also write $A_0 \xrightarrow{D} t_k$ and abbreviate $D = \langle \rho_1, \ldots, \rho_k \rangle$. The above derivation is called *complete* if $t_k \in \mathcal{T}(\mathcal{F}, \mathcal{K} \cup \{\varepsilon\})$ (*i.e.*, is contains no more non-terminals).

The *weight* of D is defined by:

$$weight(D) = \sum_{i=1}^{k} weight(\rho_i).$$

The weight of a tree $t \in \mathcal{T}(\mathcal{F}, \mathcal{K})$ *wrt* $\mathcal{G}$, also called *cost of readability* of t is:

$$cr_{\mathcal{G}}(t) = \min_{A_0 \xrightarrow{D} t} weight(D).$$

The grammar $\mathcal{G}$ might be omitted when clear from the context. The purpose of tree grammars is to define a restricted prior language of trees that are an acceptable output of transcription.

4.2 Cost of Alignment

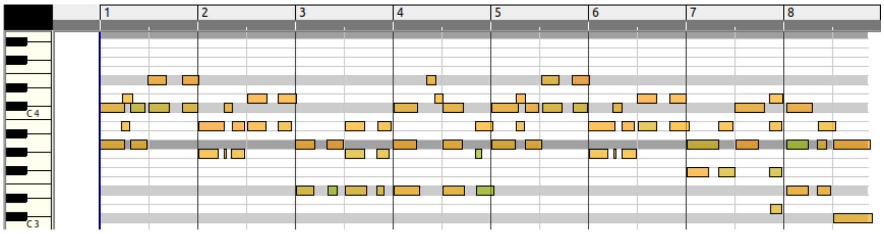
We assume that a *cost alignment function* $ca(\theta, T, \tau, \tau^{prev})$ is associated to a token T, a token type $\theta \in \mathcal{K}$, and two time points τ and τ^{prev} such that $\tau^{prev} < \tau$ or $\tau^{prev} = \perp$ with a undefined symbol $\perp$. This value is a summation of the time shifts induced by aligning all the MIDI events in the token T to the time point τ, if the token type is θ. Here, τ^{prev} is the time value to which the former token of T is aligned. If the type of T is not θ, then $ca(\theta, T, \tau, \tau^{prev}) = +\infty$. See Appenix A for the concrete defintion of a cost alignment function.
Let us consider a tree $t \in \mathcal{T}(\mathcal{F}, \mathcal{K})$, a time interval I, and the associated sequence of leaves $yield(t) = \theta_0, \ldots, \theta_p$ and sequence of pairs of token and time point $tokenize(t, I) = \langle (T_0, \tau_0), \ldots, (T_p, \tau_p) \rangle$.
The *cost of alignment* of E to the tree t and the interval I, denoted by $ca_E(t, I)$, is defined as the sum of the $ca(\theta_i, T_i, \tau_i, \tau'_i)$ for $0 \leq i \leq p$ where $\tau'_i = \perp$ if $i = 0$; otherwise $\tau'_i = \tau_{i-1}$.

4.3 Transcription Objective

The problem of transcription can be defined using the above notions, by minimizing a combination of the two cost measures defined above: the cost of alignment and the cost of readability. More precisely, given an input MIDI sequence E and a grammar $\mathcal{G}$, we call transcription of E *wrt* $\mathcal{G}$, a tree $t \in \mathcal{T}(\mathcal{F}, \mathcal{K})$ minimizing the measure:

$$ca_E(t, [0, +\infty)) + cr_{\mathcal{G}}(t) \quad (1)$$

(a) Score: Sonate für Klavier Nr.11 A dur K.331 Mov.1



(b) The piano-roll of a human piano performance

(c) A transcription by our implementation

(d) A transcription by Logic

Figure 8. A transcription example

The trees of Section 3.3 are abstract descriptions of music scores. The elementary score elements (symbols) correspond to the names of token types (notes, rests, chords *etc* in $\mathcal{K}$) labelling leaves. The MIDI pitch of every note can be extracted from the input MIDI sequence using the definition of tokens in Section 3.4. A pitch-spelling algorithm [20] is then necessary to cast these MIDI key values to note names. Moreover, the durations are encoded in t by the symbols of $\mathcal{F}$ labeling inner nodes. Additionally to the operations on time intervals, some info about the output score can be attached to the symbols of $\mathcal{F}$. For instance, we can express in a symbol div_n whether we want the notes below this symbol to be beamed or not.

4.4 Algorithm and Implementation

We designed and implemented an algorithm for the parsing problem defined in Section 4 based on ordinary tabulation technique.

Given in input a MIDI sequence E and a weighted tree grammar, it returns a tree t minimizing (1) in Section 4.3. Figure 9 presents the parsing algorithm designed based on ordinary Dynamic Programming. It assumes that the input midi-sequence E is played with a constant tempo and converted to musical time so that one bar is identical to one second.

The algorithm deals with a set of tabulated items, which is kept by the variable C, as candidates of a parsing result. each item contains

1. a current parse tree t (or a sequence of parse trees),
2. the sum $w = ca_E(t, I) + cr_{\mathcal{G}}(t)$ of the parse-tree weight and its allignment cost (or the weight sum of the parse trees and their allignment costs),
3. the set $E|_{snd(I)}$ of unprocessed events in the interval I of the parse tree(s), and
4. the assigned time τ^{prev} of the last token in the parse tree(s).

See appendix B for the detail.

The source code of this implementation is found in URL [1], and produces the command line utilities *monopase*

[1] https://gitlab.inria.fr/qparse/qparselib

Input: Midi-sequence E, tree grammar $\mathcal{G} = \langle \mathcal{N}, A_0, \mathcal{F}, \mathcal{K}, \mathcal{R} \rangle$.

Output: The tree t obtained from $(t, w) := parse([0, \infty), \emptyset, \bot)$.

Functions:

$fst(I)$ (resp. $snd(I)$) denotes $div_2(I, 1)$ (resp. $div_2(I, 2)$).

$E|_I = \{e \in E \mid \mathsf{ts}(e) \in I)\}$ is the subsequence of events in an interval I.

$parse(I, F, \tau)$:

If $F \cup E|_I = \emptyset$ **then Return** $\langle \varepsilon, 0 \rangle$ **else**

$C := parseRec(unit(I), A_0, F, \tau)$

$C' := \{ \ \langle \mathsf{bar}(t, t'), w + w' \rangle \mid \langle t, w, F', \tau' \rangle \in C, \langle t', w' \rangle := parse(rem(I), F', \tau') \ \}$

Return $\langle t, w \rangle$ with minimum weight w among $\langle t, w \rangle \in C'$.

$parseRec(I, A, F, \tau)$:

$C := \emptyset$ and $T := F \cup E|_{fst(I)}$.

For each rule $A \xrightarrow{w} u \in \mathcal{R}$, **do**

case $u = \varepsilon$**: If** $T = \emptyset$ **then** $C := C \cup \{ \langle \varepsilon, w, E|_{snd(I)}, \tau \rangle \}$

case $u = \theta \in \mathcal{K}$**:**

$C := C \cup \{ \langle \theta, w + ca(\theta, T, start(I), \tau), E|_{snd(I)}, start(I) \rangle \}$

case $u = \mathsf{div}_k(A_1 \cdots A_k)$**: Let** $C_0 = \{ \langle \langle \rangle, w, F, \tau \rangle \}$.

For $i = 1, \ldots, k$, **do**

$$C_i := \min'\Big(\bigcup_{\substack{\langle \langle t_1, \ldots, t_{i-1} \rangle, w_{i-1}, F_{i-1}, \tau_{i-1} \rangle \in C_{i-1} \\ \langle t_i, w_i, F_i, \tau_i \rangle \in parseRec(div_k(I, i), A_i, F_{i-1}, \tau_{i-1})}} \langle \langle t_1, \ldots, t_i \rangle, w_{i-1} + w_i, F_i, \tau_i \rangle \Big)$$

$C := C \cup \{ \langle \mathsf{div}_k(t_1, \ldots, t_k), w', F', \tau' \rangle \mid \langle \langle t_1, \ldots, t_k \rangle, w', F', \tau' \rangle \in C_k \}$.

Return C

$\min'(C)$: The set of $\langle S, w, F, \tau \rangle \in C$ having minimum w for each $\langle F, \tau \rangle$.

Figure 9. Parsing algorithm

for parsing monophonic/homophonic input in branch mono2 and *drumparse* for drums in branch beta. The current implementation recognizes the token type of partial continuation as an ordinary continuation.

4.5 Example

In addition to former examples of monophonic [2] and drum [3] transcription, we present an example of homophonic transcription from a MIDI [4] in Figure 8. It is difficult to say that the result Figure 8(c) is a success, but it gives us valuable hints for future work.

1. Some extremely short notes (see the piano roll in Figure 8(b)) are recognized as grace notes.
2. There exist in the output mei unnecessary printed accidentals, which are removed manually to ease readability in Figure 8 (b). Such hander is not yet implemented in current monoparse.
3. The generation of scores in musicXML format has a bug that improperly transforms the continuations of chords. The following notes seem strange: the 2nd note in the 4th bar, the 2nd and 4th note in the 7th bar, and the 2nd note in the last bar.

[2] https://qparse.gitlabpages.inria.fr/docs/examples

[3] https://gitlab.inria.fr/transcription/gmdscores

[4] https://www.trs.css.i.nagoya-u.ac.jp/~sakai/tenor2024/

5. CONCLUDING REMARKS

We have proposed an approach for the tokenization of MIDI sequences which supports score elements such as rests and ornaments. Implemented as a Dynamic Programming algorithm, it is integrated into a framework of transcription by parsing.

The identification of grace-chords, as well as arpeggiated chords, shall be added to the cases of Section 2.4. Another future objective is the processing of polyphonic input, like piano MIDI files. When all voices are mixed into the same MIDI file, the application of a voice separation procedure is required. An important question is then whether voice separation should be performed before, after, or jointly to the parsing described in Section 4. Note that in the first case, voice separation would have to deal with unquantized MIDI input, and with quantized rhythms in the two latter cases.

Acknowledgments

This research was supported by JSPS Kaken 20H04302.

6. REFERENCES

[1] D. Grune and C. J. Jacobs, *Parsing Techniques*, ser. Monographs in Computer Science. Springer, 2008 2nd edition.

[2] M. Zeng, X. Tan, R. Wang, Z. Ju, T. Qin, and T.-Y. Liu, "MusicBERT: Symbolic music understanding with large-scale pre-training," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Association for Computational Linguistics, 2021, pp. 791–800.

[3] Y. Ren, J. He, X. Tan, T. Qin, Z. Zhao, and T.-Y. Liu, "Popmag: Pop music accompaniment generation," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1198–1206.

[4] G. Hadjeres and L. Crestel, "The piano inpainting application," arXiv preprint arXiv:2107.05944, 2021.

[5] S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, "This time with feeling: Learning expressive musical performance," *Neural Computing and Applications*, vol. 32, pp. 955–967, 2020.

[6] Y.-S. Huang and Y.-H. Yang, "Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 1180–1188.

[7] W.-Y. Hsiao, J.-Y. Liu, Y.-C. Yeh, and Y.-H. Yang, "Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, 2021, pp. 178–186.

[8] J. Gardner, I. Simon, E. Manilow, C. Hawthorne, and J. Engel, "MT3: Multi-task multitrack music transcription," *arXiv preprint arXiv:2111.03017*, 2021.

[9] N. Fradet, J.-P. Briot, F. Chhel, A. E. F. Seghrouchni, and N. Gutowski, "Byte pair encoding for symbolic music," 2023. [Online]. Available: https://arxiv.org/abs/2301.11975

[10] J. Ens and P. Pasquier, "MMM : Exploring conditional multi-track music generation with the transformer," 2020.

[11] N. Fradet, J.-P. Briot, F. Chhel, A. El Fallah Seghrouchni, and N. Gutowski, "MidiTok: A python package for MIDI file tokenization," in *22nd International Society for Music Information Retrieval Conference (ISMIR Late-Breaking Demo)*, 2021.

[12] F. Foscarin, F. Jacquemard, P. Rigaux, and M. Sakai, "A Parse-based Framework for Coupled Rhythm Quantization and Score Structuring," in *Mathematics and Computation in Music (MCM)*, ser. Lecture Notes in Artificial Intelligence, vol. 11502. Springer, 2019.

[13] K. Shibata, E. Nakamura, and K. Yoshii, "Non-local musical statistics as guides for audio-to-score piano transcription," *Information Sciences*, vol. 566, pp. 262–280, 2021.

[14] M. Suzuki, "Score transformer: Transcribing quantized midi into comprehensive musical score," in *Proceediongs of the International Soc. Music Information Retrieval Conference (ISMIR, Late-breaking Demo)*, 2021.

[15] C. Agon, G. Assayag, J. Fineberg, and C. Rueda, "Kant: a critique of pure quantification," in *International Computer Music Conference Proceedings (ICMC)*, 1994, pp. 52–59.

[16] A. T. Cemgil, B. Kappen, and P. Desain, "Rhythm quantization for transcription," *Computer Music Journal*, vol. 24, no. 2, pp. 60–76, Jul. 2000. [Online]. Available: http://dx.doi.org/10.1162/014892600559218

[17] F. Jacquemard and L. Rodriguez de la nava, "Symbolic weighted language models, quantitative parsing and automated music transcription," in *26th International Conference on Implementation and Application of Automata (CIAA)*, 2022.

[18] M. Digard, F. Jacquemard, and L. Rodriguez de la Nava, "Automated transcription of electronic drumkits," in *4th International Workshop on Reading Music Systems (WoRMS)*, 2022.

[19] J. Yust, *Organized Time*. Oxford University Press, 2018.

[20] D. Meredith, "The PS13 pitch spelling algorithm," *Journal of New Music Research*, vol. 35, no. 2, pp. 121–159, 2006.

A. COST ALIGHNMENT FUNCTION

Let E be an input MIDI sequence. For a token type $\theta \in \mathcal{K}$ and a token T of E, let $\theta(T)$ be defined by:

$\theta(T) = 1$	if T has type θ and θ is valid in the case of input considered (see Section 2.5),
$\theta(T) = +\infty$	otherwise,

The cost alignment function used in the implementation is as follows:

$$ca(T, \theta, \tau, \tau_p) = \frac{\theta(T)}{\Delta(\tau_p, \tau)} \sum_{e \in T} \alpha_\sigma\big(r(e, T)\big) \cdot |\mathsf{ts}(e) - \tau|$$

where $\Delta(\tau^{prev}, \tau) = \tau - \tau^{prev}$ if $\tau^{prev} \neq \perp$; otherwise a fixed constant value. σ is the sign of $\mathsf{ts}(e), \tau$ and α_+ and α_- associate appropriate constants to the role $r(e, T)$ of e in T. Typically, $\alpha_{\mathsf{offL}} > \alpha_{\mathsf{offR}}$ because note-off often occurs earlier than expected.

B. ALGORITHM

This section gives details of the algorithm. Each function works as follows.

parse: Assuming the following arguments:

- an interval I of midi-sequence E to be parsed,
- a set F of events in prior to $start(I)$, and
- the time point τ aligned the last parsed token before $start(I)$,

it returns a pair of the best weighted tree t and its weight w for midi-sequence $F \cup E$ with alligning all events in F to $start(I)$.

parseRec: calculates possible best trees within one bar. Assuming the following arguments:

- an interval I of midi-sequence E to be parsed,
- a non-terminal symbol A of the grammar,
- a set F of events in prior to $start(I)$, and
- the time point τ aligned the last parsed token.

it returns a set of tuples $\langle t, w, F', \tau' \rangle$ consisting of the tree t with weight w and root nonterminal symbol A for each possble pair of F' and τ' determined in the calculation.

ENGRAVING ORIENTED JOINT ESTIMATION OF PITCH SPELLING AND LOCAL AND GLOBAL KEYS

Augustin Bouquillard
Université Paris Dauphine, PSL
augustinus.bouquillard@orange.fr

Florent Jacquemard
Inria Paris and Cedric/CNAM
florent.jacquemard@inria.fr

ABSTRACT

We revisit the problems of pitch spelling and tonality guessing with a new algorithm for their joint estimation from a MIDI file including information about the measure boundaries. Our algorithm does not only identify a global key but also local ones all along the analyzed piece. It uses Dynamic Programming techniques to search for an optimal spelling in term, roughly, of the number of accidental symbols that would be displayed in the engraved score. The evaluation of this number is coupled with an estimation of the global key and some local keys, one for each measure. Each of the three informations is used for the estimation of the other, in a multi-steps procedure.

An evaluation conducted on a monophonic and a piano dataset, comprising 216 464 notes in total, shows a high degree of accuracy, both for pitch spelling (99.5% on average on the Bach corpus and 98.2% on the whole dataset) and global key signature estimation (93.0% on average, 95.58% on the piano dataset).

Designed originally as a backend tool in a music transcription framework, this method should also be useful in other tasks related to music notation processing.

1. INTRODUCTION

In symbolic music representations, pitches are expressed in different ways. In their simplest form, in the MIDI standard, they are encoded as integers corresponding to a number of keys on a device. The representation of pitches is much more involved in Common Western Music Notation (CWN), where the denotation of each note depends on the musical context of occurrence: the tonality (key) of the piece, the voice-leading structure (ascending or descending melodic movements), the harmonic context...

If we reason modulo octaves, every pitch class, between 0 and 11 semitons, can be denoted in several ways, using a note name, in C, D, E, F, G, A, B, and an accidental mark in 𝄫, ♭, ♮, ♯, 𝄪, acting as a pitch class modifier. *Pitch-Spelling* (PS) is the problem of choosing appropriate names to denote some given MIDI pitch values in CWN.

In the tonal system, fixing a (global) key for a piece defines some default, privileged, names and accidentals. This rule serves two important purposes in practice for the reader of a music score. On the one hand, since the default accidentals are not printed, the number of symbols displayed on the score is reduced and hence the readability is eased. On the other hand, the key signature immediately poses a tonal context for the piece, and the presence of other (non default) accidentals constitutes an indication of the tonal function of the notes, which provides insight into the composer's intention, in particular regarding local key changes. For instance, let us consider the last chord of the 56th measure of Mozart's c minor Sonata's 1st movement (Figure 1a), which is resolved in the next bar on the second inversion of an E♭ Major triad, comprising a G natural in the first voice of the left hand. By spelling the highest note of the left hand in the chord of interest as G♭ instead of F♯, Mozart chooses not to comply with the principle of selecting an ascending accidental when a voice goes up, and rather to express the harmonic function of dominant of the dominant in E♭, which is the new tonality he is heading to at this moment. In the same movement, measure 125 (Figure 1b), a chord composed of the exact same notes is then spelled differently: an F♯ has replaced the G♭, as the tonal context is shifting back to the main tone and the harmonic function has changed to dominant of the dominant in c minor this time. This constitutes one of the numerous examples (*e.g.*, [1], Chopin's first Ballade or Tristan's chord) of one chord being spelled differently depending on its harmonic nature or tonal function and in this case regardless of the melodic movements of its voices, which shows that pitch spelling is indeed revealing in terms of creative intent and not only useful to the readability of a piece.

There is therefore a strong interdependency between the problems of *Pitch-Spelling*, and (local and global) *Key Estimation* (KE).

A large variety of PS algorithms have been proposed, which are able to guess the spelling of reference corpora with a high degree of accuracy. Several such algorithms have been designed according to musicological criteria, selecting note spellings based on the analysis of voice-leading, interval relationships and local keys [2, 3, 4], and a principle of parsimony (minimisation of the number of accidentals) [5]. Some procedures, also based on musicological intuitions, reduce PS to optimisation problems in appropriate data structures, *e.g.*, the Euler lattice [6] or weighted oriented graphs [7]. In other approches, datasets of music scores are used to train models such as HMM [8] or RNN [9] for PS. The system in the latter reference, esti-

(a) Mozart, Sonata no 14 in c minor, mvt 1, measures 56-57.

(b) Mozart, Sonata no 14 in c minor, mvt 1, measures 125-126.

Figure 1. Tonal cues in chord spellings in Mozart's Sonata no 14.

mates, in addition to PS, a key signature for the input piece, with a high accuracy.

In this paper, sharing with [9] the goal of joint Pitch Speling and Key Estimation from MIDI data, we generalise the KE problem from global key signature to global and local keys. Our approach is algorithmic, following the very simple and intuitive idea to minimise the number of accidental symbols *as they would appear on the engraved score*. Similar criteria of parsimony are also found *e.g.*, in [5]. A difference is that we are applying this principle in a tonal context: assuming a key for the piece to spell, we count the accidentals for the possible spellings according to the *rules of engraving*, *i.e.*, considering that the accidentals defined by the key signature are not printed, and that a printed accidental holds for the embedding measure.

Our procedure works in two steps. Assuming that the MIDI input is divided into measures, we first estimate, for each key in a given set, the least number of printed accidentals for all possible spellings in all measures. We use this information in order to estimate a local key for each measure, and each possible global key. Then, in a second step, the above evaluation of the quality of spellings is refined by taking into account (in addition to the number of printed accidentals) the proximity of spellings to the evaluated local keys. The estimated global key is the one giving the best evaluation (cumulated for all measures), and the spelling chosen is the one computed for this global key.

The prior division into measures is crucial in our approach because of its role in the rules of engraving (cf Section 3.1). This assumption is not required in the above cited papers, either those following algorithmic or ML approaches. The algorithmic papers rather use a sliding window of parametric size. With that respect, our procedure applies to more restricted cases. However, this assumption makes sense when dealing with quantized MIDI data, in particular in the last step of a music transcription framework, after rhythm quantization. Also, the estimated global and local keys, which are somehow a side effect of our PS procedure, may be useful, as descriptors, in other tasks of processing of MIDI data for which metrical information is known. Finally, apart from the boundaries of measures, we do not need other information such as note durations or voice separation (see the discussion on that point in Section 5).

To summarize our contributions, we propose an original and somewhat naive approach to two old problems, which combines them and obtains very good results on challenging datasets. In Section 2, we present the preliminary notions used to state the problems of PS and KE. The reader well versed into these problems may skip this section. We detail our procedure in Section 3 and present in Section 4 its evaluation on two datasets (one monophonic and one piano), for a total of 216 464 notes, which has given very good results.

2. PRELIMINARIES

We call *part* a polyphonic sequence of notes, organized in measures (bars). Typically, it shall correspond to one staff in CWN. Every note is given by values of pitch, onset and duration. The representation of these values is described in the two next subsections, before another subsection treating the subject of keys and signatures.

2.1 Pitch Representations

There are two classical alternative representations for the note *pitch*, distinguishing the input and output of a pitch-spelling algorithm:

- a *MIDI* value, in 0..128, which is a number of semitones [10],
- a *spelling*, made of:
  - a note *name* in A, . . . , G,
  - a symbol of *accidental*, amongst ♮ (natural), ♭ (flat), 𝄫 (double flat), ♯ (sharp), 𝄪 (double sharp),
  - an octave number in -2. . . 9.

The lowest MIDI value 0 corresponds to C♮-1 (B♯-2), and the highest one, 128, is G♯9 (A♭9). The 88 keys of a piano correspond to the MIDI numbers 21 (A♮0) to 96 (C♮7). A MIDI value modulo 12, is called *pitch class*. The pitch class of a note ν is denoted by $pc(\nu)$.
Every note name is associated a unique pitch class: 0 for C to 11 for B. An accidental symbol acts as a pitch class modifier: 𝄫, ♭, ♮, ♯, 𝄪, respectively add -2, -1, 0, 1, and 2 to the pitch class of the note name component in a spelling. In the following, the symbol ♮ is sometimes omitted, *i.e.*, written as a space. Altogether, with the octave component, this principle permits to associate a unique MIDI value to a given spelling. In the other direction, there are several alternative valid spellings for a given MIDI value, as summarized in Figure 2 for the 12 pitch classes.

For every pitch class, the 2 or 3 alternative spellings of Figure 2 have different names. Hence, for the purpose of finding a spelling for a pitch of MIDI value m, is is sufficient to choose one of the 2 or 3 possible names for m modulo 12. The corresponding accidental symbol and octave number can then be deduced from the name chosen.

pitch class	spelling$_1$	spelling$_2$	spelling$_3$
0	D𝄫	C	B♯
1	D♭	C♯	[B𝄪]
2	E𝄫	D	C𝄪
3	[F𝄫]	E♭	D♯
4	F♭	E	D𝄪
5	G𝄫	F	E♯
6	G♭	F♯	[E𝄪]
7	A𝄫	G	F𝄪
8	A♭	G♯	
9	B𝄫	A	G𝄪
10	[C𝄫]	B♭	A♯
11	C♭	B	A𝄪

Figure 2. Enharmonic spellings for each pitch class.

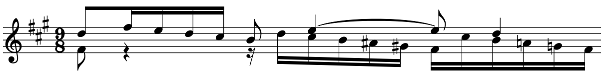


Figure 3. Bach, Fugue in A Major BWV864, measure 33, rh.

Example 2.1. Figure 3 presents the right-hand part in measure 33 of the Fugue in A Major BWV864 of J.S. Bach. The MIDI values of the two notes on the first beat of the measure are 66, with possible spellings either F♯4 or G♭4, and 74, with possible spellings either D5, or C𝄪5, or E𝄫5. Here the chosen spelling does not induce any accidental, as F♯ is included in the key signature. An alternative spelling to the one on the score for these two notes could thus be G♭, D but it would generate an added accidental on the G since the signature does not contain any G♭; hence we observe the importance of the key signature for pitch spelling.

2.2 Time information

In this paper, we assume that the *onset* and *duration* values of notes are rational numbers, expressed in fraction of a measure. The choice of a time unit is not really relevant in this work. The informations related to time that is important in our procedure are actually:

- an ordering and equality relation on onsets, for sorting the notes in input and detecting notes starting simultaneously,
- the number of measure to which a note belongs, *i.e.*, the detection of bar changes in the flow of input notes.

We call *grace-note* a note of theoretical duration zero. We use this term generically for a note which is part of an ornament (appoggiatura, gruppetto, mordent, trill *etc*). Two notes are called *simultaneous* if they have the same onset and are not grace notes. This definition can correspond to notes involved in the same chord, or notes in different voices and starting simultaneously.

Example 2.2. For instance, the two first notes of Figure 3 do not really constitute a chord but they are *simultaneous* since they share the same onset: 0. The time signature for the measure is 9/8, hence the two first notes F♯ and D♮ both have a duration of $\frac{1}{9}$ bar, whereas the next semi-quaver F♯5, starting at the onset $\frac{2}{9}$, has a duration of $\frac{1}{18}$.

2.3 Key Signature and Key

A Key Signature (KS) is an integral number k between -7 and 7, which indicates that by default, $|k|$ note names shall be altered by a sharp accidental symbol (when $0 < k \leq 7$) or a flat accidental (when $-7 \leq k < 0$). The names of the notes altered are defined according to the order of fifths: F♯, C♯, G♯, D♯, A♯, E♯, B♯ for the sharps (k positive) and B♭, E♭, A♭, D♭, G♭, C♭, F♭, for the flats (k negative). In a score, a KS indication is placed at the beginning of a part and will influence the display of the spelling of notes in every measure, as explained in Section 3.1. The KS can be changed during a part.

From the notational point of view (*i.e.*, for engraving), the choice of KS will change drastically the way the notes are displayed on the score, as it influences the number of accidental symbols printed, hence the readability of the score.

From a musical point of view, a KS together with a *mode*, defines a *global key* K for a piece, which is a central notion in tonal music. Indeed, the key identifies a diatonic scale, whose first note (amongst seven notes), called the *tonic* note, represents the main tonal focus of the piece. In Figure 4, we describe 15 key signatures and the tonic of associated keys. Keys defined by the same KS and different modes are called *relative*.

Additionally to the global key of a piece (or part), some alternative *local keys* can be identified for extracts of the piece, and might diverge from the global key through *modulations*. In this work, we shall estimate a local key for each measure in a part, and this information is used for pitch spelling (Section 3.3).

We shall consider below, in our joint Pitch Spelling and Key Estimation procedure, the *major* mode, and three minor modes: *harmonic minor*, *melodic minor* (also called *ascending minor*) and *natural minor* (also called *descending minor* or *Aeolian*). The three minor modes differ by the alteration of some degrees in the scale, notably the leading tone (or subtonic in the case of the natural minor scale) and the submediant with accidentals not in the key signatures: seventh degree (raised) for *harmonic minor*, and sixth and seventh degrees (raised) for *melodic minor*.

Only the major and harmonic minor modes are used for global Key Estimation. The melodic and natural minor modes are only used for local Key Estimation, see Section 3.3 and Section 4.3 for a discussion on occurrences of these modes in pieces used for evaluation. We shall consider a measure of distance between keys defined by Gottfried Weber in [11] (see Appendix A.1), see also [12] for other use of Weber's table in the context of Key Estimation.

In theory, the list of key signatures can be extended on the right and on the left, respectively through double sharps and double flats. For instance, with $k = 8$ (G♯ major), F is altered with a double sharps (F𝄪), with $k = 9$ (D♯ major), F and C are altered with double sharps (F𝄪, C𝄪), *etc*. We do not consider the case of extended KS in this work, as they are very rarely found.

key signature	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7
major mode	C♭	G♭	D♭	A♭	E♭	B♭	F	C	G	D	A	E	B	F♯	C♯
minor modes	A♭	E♭	B♭	F	C	G	D	A	E	B	F♯	C♯	G♯	D♯	A♯
C	♭	♭								♯	♯	♯	♯	♯	♯
D	♭	♭	♭	♭								♯	♯	♯	♯
E	♭	♭	♭	♭	♭	♭								♯	♯
s_0 : F →	♭								♯	♯	♯	♯	♯	♯	♯
G	♭	♭	♭								♯	♯	♯	♯	♯
A	♭	♭	♭	♭	♭								♯	♯	♯
B	♭	♭	♭	♭	♭	♭	♭								♯

Figure 4. Key signatures and keys.

In both major and minor modes, the keys associated with key signatures respectively -7 and 5, -5 and 7, and -6 and 6 have tonics with the same pitch class, but different names. These keys are called *enharmonic*. They correspond to the 3 first and 3 last columns in Figure 4. Melodies written in either of two enharmonic keys cannot be distinguished by ear (in equal temperaments) as changing a key for its enharmonic preserves not only the intervals but also the pitch class of every note. Therefore, spelling in one or the other of enharmonic keys is essentially a matter of choice. Usually, the keys with KS 5 (5 sharps, *e.g.*, B major) or KS -5 (5 flats, *e.g.*, D♭ major) are preferred over their enharmonic equivalents with respectively KS -7 (*e.g.*, C♭ major) and KS 7 (*e.g.*, C♯ major). But it is not always the case, for instance the Prelude and Fugue, BWV 848, of Bach's Well-Tempered Clavier are in C♯ major.

3. ALGORITHMS

We present in this section the procedure that we are using to estimate, in the same process, best spellings, a global key and one local key per measure, for a given sequence of input notes. Roughly, it works measure by measure, by comparing the number of accidentals in all spellings for all candidate global keys in a given set. This exhaustive comparison is done through dynamic programming techniques, based on the principle used to decide which accidentals must be printed or not in CWN (presented in next Section 3.1). The estimated local keys are used to refine the selection of spellings in each measure, in case of ties.

3.1 Notational Convention Modulo

In order to lighten the notations, and ease the readability, some accidental symbols are not printed in scores in CWN. Following a principle of parsimony, the notational conventions are roughly that: the accidentals in the key signature are omitted by default, and the other accidentals need not be repeated in the same measure. There is moreover a restriction to this rule, summarised as follows in [13]: *It applies only to the pitch at which it is written: each additional octave requires a further accidental.*

Let us present below a relaxed version of this convention, without the above restriction for octaves. Our approach amounts in some way to reasoning modulo 12, replacing the notion of pitch by the notion of pitch class. This application of the principe of parsimony is more suitable to the problem of Pitch Spelling [5], which is more concerned in *counting* the accidentals than in *printing* them.

Formally, the definition of the relaxed convention is based on a notion of *spelling state*, or state for short. Such a state s is a mapping from the note names in $\{\mathsf{A}, \ldots, \mathsf{G}\}$ into accidental symbols in {𝄫, ♭, ♮, ♯, 𝄪}. Let us assume a key signature $k \in \{-7, \ldots, 7\}$ and a measure containing a sequence $\nu_1, \ldots, \nu_p$ of p notes, enumerated by increasing onsets. An initial state s_0 for the measure is built from k as expected (see Figure 4): for $k = 0$, $s_0(n) = ♮$ for all name n, for $k = 1$, $s_0(\mathsf{F}) = ♯$ and $s_0(n) = ♮$ for all other name n, *etc.* For $1 \leq i \leq p$, the state s_i is computed by updating the previous state s_{i-1} as follows, where n and a are the name and accidental of ν_i:

$i.$ if $s_{i-1}(n) = a$, then $s_i = s_{i-1}$, and a is *not printed,*

$ii.$ otherwise, $s_i(n) = a$, $s_i(n') = s_{i-1}(n')$ for all $n' \neq n$, and a is *printed.*

Example 3.1. In Figure 3, for instance, at all onsets before $\frac{5}{9}$, the spelling state is composed of F♯, C♯, G♯ with every other note of the scale being natural. On onset $\frac{5}{9}$ however, the state changes for the first time in the measure and now contains A♯ instead of A♮.

The next onset also induces a change in the state with G♮ being replaced by G♯, which is a note present in the ascending minor melodic mode of B. It is interesting to remark that Bach used it in a descending motion, therefore a pitch spelling process relying too much on motion direction between notes would have failed here.

The spelling state then stays the same until onset $\frac{15}{18}$ with the return of A♮ (belonging to the natural minor mode of B) and finally G♮ at onset $\frac{16}{18}$, hence the last states of the measure are the same as the one it started with.

3.2 Processing One Measure in a Key

Let K be a key with key signature $k \in \{-7 \ldots 7\}$ and let $\bar{\nu} = \nu_1, \ldots, \nu_\ell$ be a sequence of notes inside a measure, with known MIDI values and unknown spellings. In order to find the spellings for $\bar{\nu}$ with the least number of printed accidentals, according to the conventions of Section 3.1, it is fortunately not necessary to enumerate all the $O(3^\ell)$ possible spellings of $\bar{\nu}$. Indeed, for that purpose, it is sufficient to perform a shortest path search, using the state structure of Section 3.1. Before presenting the details of the search method, let us formulate an additional hypothesis, that is

Figure 5. Beethoven, Sonata 21 "Waldstein", measures 1-3, lh.

relevant musically and turned out to be important from a combinatoric point of view:

(h) *two simultaneous notes in the same pitch class must have the same name.*

The hypothesis (*h*) is not a strict notational convention, although counter examples are very rare. However, we do not assume the other direction (simultaneous notes with same name must be in the same pitch class), as it can occur fairly frequently in tonal music, at least from the end of the nineteenth century, for instance in a dominant ninth chord with an appoggiatura on the ninth (*e.g.*, D F♯ A C F♮ in G major), much appreciated by Ravel, among others.

Example 3.2. Without the hypothesis (*h*) stated above, the presence of numerous chords inside of one measure made the complexity of PSE explode in cases such as the beginning of the Waldstein Sonata, displayed in Figure 5. The treatment of simultaneous notes allows the algorithm not to treat all possible spellings for each apparition of doubled notes inside a chord (in this example one of the C's in the repeated C major chord need not be treated once the other C's spelling has been determined).

In order to ensure the hypothesis (*h*), we consider another state variable $c \in \mathcal{C} = \{\mathsf{A}, \ldots, \mathsf{G}\}^{\{0,\ldots,11\}}$ which is a partial mapping of pitch classes into note names. It is used to memorize the name associated to a pitch class when processing a subsequence of simultaneous notes in $\bar{\nu}$. The spelling of a note with pitch class p and note name n is called *compatible* with c iff $c(p)$, if defined, is equal to n.

Let $\mathcal{S} = \{𝄫, ♭, ♮, ♯, 𝄪\}^{\{\mathsf{A},\ldots,\mathsf{G}\}}$ be the set of possible values for the state variable s defined in Section 3.1, $I = \{1 \leq i < \ell \mid \nu_i \text{ and } \nu_{i+1} \text{ are simultaneous}\}$, $\bar{I} = \{1, \ldots, \ell\} \setminus I$, and let us consider the following set of configurations:

$$V = \big\{\langle s_0, 0\rangle\big\} \cup \mathcal{S} \times \bar{I} \cup \mathcal{S} \times \mathcal{S} \times \mathcal{C} \times I.$$

The configuration $\langle s_0, 0\rangle$ is initial, where s_0 is the initial state associated to K as in Section 3.1. In configurations of the form $\langle s, i\rangle \in \mathcal{S} \times \bar{I}$ we have one state s for a note index $i \in \bar{I}$. These configurations are dedicated to the processing of single notes. The other configurations, of the form $\langle s, t, c, i\rangle \in \mathcal{S} \times \mathcal{S} \times \mathcal{C} \times I$ contain additional information in t and c for processing a sub-sequence of simultaneous notes, ensuring in particular hypothesis (*h*).

We consider a set of transitions $E \subset V \times \mathbb{N} \times V$, containing the *weighted edges* of one of the following forms:

$$\langle s, i-1\rangle \xrightarrow[n,a]{w} \langle s', i\rangle \quad (1)$$

$$\langle s, i-1\rangle \xrightarrow[n,a]{w} \langle s', s, c', i\rangle \quad (2)$$

$$\langle s, t, c, i-1\rangle \xrightarrow[n,a]{w} \langle s', t, c', i\rangle \quad (3)$$

$$\langle s, t, c, i-1\rangle \xrightarrow[n,a]{w} \langle s', i\rangle \quad (4)$$

Every above case means that there exists a spelling of ν_i with name n and accidental a, which is moreover compatible with c in cases (3) and (4). In transitions (1)-(4), s' is the update of s as defined in Section 3.1 (cases i, ii). However, since the purpose of state s is no more score engraving like in Section 3.1, but the the search of a pitch spelling, we shall distinguish below, for the definition of the weight value w, the cases when the accidental a is *counted* from the cases when it is *not counted* (and not anymore *printed* or *not printed* like in Section 3.1).

The transition (1) processes the single note ν_i which is not simultaneous with ν_{i+1}. The condition for counting a is the same as the condition for *printing* in Section 3.1 (cases i, ii):

- if $s(n) = a$, then $s' = s$ and a is *not counted*,
- otherwise, $s'(n) = a$, $s'(n') = s(n')$ for all $n' \neq n$, and a is *counted*.

The transition (2) initiates the processing of a subsequence of two or more simultaneous notes, when ν_i is simultaneous with ν_{i+1}. It makes a copy of the current state s in the second component, and create $c' = \{\langle p, n\rangle\}$, with $p = pc(\nu_i)$. The accidental a is *counted* or *not*, and s updated into s', under the same conditions as in (1).

The transition (3) performs one step of the processing of a subsequence of simultaneous notes, again when ν_i is simultaneous with ν_{i+1}. It propagates the frozen copy of state t (without modifying it) and updates c into c' as follows:

- if $c(p)$ is defined and equal to n, then $c' = c$, $s' = s$ and a is *not counted*,
- if $c(p)$ is undefined, then $c' = c \cup \{\langle p, n\rangle\}$, moreover, if $t(n) \neq a$, s is updated into s' with $\langle n, a\rangle$ and a is *counted*, otherwise, $s' = s$ and a is *not counted*,

The transition (4) terminates the processing of a subsequence of simultaneous notes, when ν_i is not simultaneous with ν_{i+1}. The conditions for counting a and updating s into s' are the same as in (1).

Finally, the weight value of each transition in (1)-(4) is:

1. $w = 0$ if a is *not counted*, or if K is in *harmonic minor* or *melodic minor* mode, and ν_i corresponds to an altered (raised) degree in the corresponding scale and a is the accidental for this degree in the scale.
2. otherwise, $w = 1$ if $a \in \{♭, ♮, ♯\}$, and $w = 2$ if $a \in \{𝄫, 𝄪\}$.

The purpose of the above exception for minor modes is to help the estimation of keys with minor modes in the next section.

The oriented graph $\mathcal{G} = \langle V, E \rangle$ is acyclic. The vertex $\langle s_0, 0 \rangle$ is called the *source* of $\mathcal{G}$ (it has no incoming edge) and the vertices of $\mathcal{S} \times \{\ell\}$ are called *targets* of $\mathcal{G}$ (they have no outgoing edge).

Intuitively, a path starting from the source vertex $\langle s_0, 0 \rangle$ and ending with a target vertex of V, and following the edges of E, describes a spelling of the sequence of notes $\bar{\nu}$ in input. The cumulated sum of the weight values w of edges involved in such a path amounts to a count of the accidental symbols printed.

The problem of searching for spellings of the notes of $\bar{\nu}$ with a minimal number of printed accidentals reduces to the search in $\mathcal{G}$ of a path with a minimal cumulated weight, from the source vertex into a target vertex of V. That can be done in time $O(|V| + |E|)$ with a greedy Viterbi algorithm [14], tagging the vertices of V with cumulated weight values. It proceeds in a forward way, starting from the source vertex, such that the tag of a vertex v is the weight of a minimal path leading to v.

For efficiency, the graph $\mathcal{G}$ is built on-the-fly, starting from the source vertex $\langle s_0, 0 \rangle$, and adding vertices along with the computation of their tagging, following edges satisfying the above conditions. This strategy ensures a pruning of unnecessary search branches, making the approach efficient for a use on practical cases, as described in Section 4.

3.3 Processing One Part

We assume given a sequence of m measures containing notes with known MIDI values and unknown spellings. Let us consider a fixed array of keys $\bar{K} = K_1, \ldots, K_n$. Our Pitch Spelling approach works by filling a table of dimensions $n \times m$ with best spellings for every key in $\bar{K}$ and every measure, using variants of the algorithm of Section 3.2 for each cell. Then, one estimated global key is selected in $\bar{K}$, according to the table content, and the spellings in the corresponding row of the table can be applied to the input.

We proceed in several steps, building actually several tables, with several variants for the domain of weight values.

3.3.1 Step 1: Computation of a First Spelling Table.

In the first step, we compute a table T of dimension $n \times m$, such that the cell $T[i, j]$ contains the cumulated weight (in $\mathbb{N}$) of a best spelling, according to the algorithm of Section 3.2, with $K = K_i$ and $\bar{\nu}$ are the notes of measure j.

3.3.2 Step 2: Estimation of Candidate Global Keys

We compute the sum of each row in T, and save in a list L of candidate global keys the keys of $\bar{K}$ in *major* and *harmonic or natural minor* modes with a smallest sum (there may be ties).

3.3.3 Step 3: Computation of a Grid of Local Keys

In a second table G of dimensions $n \times m$, we shall store estimated local keys. The cell $G[i, j]$ shall contain the local key estimation for the measure number j, assuming that K_i is the global key. The estimation is done column by column (*i.e.*, measure by measure).

For each measure number $0 \leq j \leq m$, we compute a ranking R^{w}_j of the keys in $\bar{K}$, according to the values (in $\mathbb{N}$) of $T[0, j], \ldots, T[n, j]$, the smallest weight value being the best.

Then, for the same j and for each $0 \leq i \leq n$ we compute two other rankings $R^{\mathsf{p}}_{j,i}$ and $R^{\mathsf{g}}_{j,i}$ of $\bar{K}$, according to respectively the Weber distance to the previous estimated local key and the assumed global key (the smallest distance is the best). More precisely, the value associated to the key $K_{i'} \in \bar{K}$, with $0 \leq i' \leq n$, for computing its rank in $R^{\mathsf{p}}_{j,i}$ is (see Appendix A.1 for the Weber table):

- the Weber distance between $K_{i'}$ and K_i if $j = 0$,
- the Weber distance between $K_{i'}$ and the previous local key in the same row $G[i, j-1]$, if $j > 0$.

And the value associated to the key $K_{i'}$ when computing its rank in $R^{\mathsf{g}}_{j,i}$ is the Weber distance between $K_{i'}$ and K_i.

Finally, for j and i, we aggregate the three rankings R^{w}_j, $R^{\mathsf{p}}_{j,i}$, and $R^{\mathsf{g}}_{j,i}$ into a unique ranking $R_{j,i}$, by comparing, for each $0 \leq i' \leq n$, the mean of its three ranks from all the rankings – see [15] about this method. The estimated local key in $G[i, j]$ is the one with the best rank in $R_{j,i}$.

In practice, the computation of G can be restricted to the rows present in the list L extracted at *step 2*.

3.3.4 Step 4: Computation of a Second Spelling Table

In this final step, we compute a last table U of dimension $n \times m$ (or *length of* $L \times m$ since the computation of U can also be restricted to the rows present in the list L), using the same algorithm as in *step 1*, but with more involved weight values.

These weight values, used to replace the $w \in \mathbb{N}$ in Section 3.2, are tuplets of integers, with the following components, for $U[i, j]$:

1. the value $w \in \mathbb{N}$ of Section 3.2,
2. the number of accidentals which do not occur in the scale associated to the estimated local key, for the notes $\bar{\nu}$ of measure j and the global key $K = K_i$,
3. the number of spellings not in the chromatic harmonic scale [1] of the estimated local key,
4. the number of accidentals with a color different from the global key signature k_i of K_i (*i.e.*, 𝄫 or ♭ when $k_i > 0$ and ♯ or 𝄪 when $k_i < 0$),
5. the number of C♭, or B♯, or F♭, or E♯.

The components (2) and (3) will second the former weight value (1) in order to refine the search of best spelling thanks to the information gained with the local key estimation in *step 3*. The two last components (4) and (5) have been added for the purpose of tie-breacking.

We consider two orderings on the domain of above weight values:

$W_{\ell ex}$: the lexicographic comparison of the tuplets with the 5 components (1)-(5),

W_+: the lexicographic comparison of the 4-uplets with, for first component the sum of (1) and (2), and for next components (3), (4), and (5) respectively.

Using the same technique as in *step 2*, we extract from L one unique estimated global key $K_i \in \bar{K}$, using the above refined weight values, and apply the spellings found in the row i of the table U.

Example 3.3. A printout of the first table (*step 1*) computed when treating the Fugue BWV 864 (see the extract in Figure 3) can be found in Appendix A.2.

Here, 2 global candidates are obtained (*step 2*): A major (3 sharps) and F♯ minor (3 sharps) as their cumulated row costs are close and inferior to all the others. Only their corresponding rows will then be computed in the second table (*step 4*), this time taking into account local tonal analysis (performed during *step 3*) to refine the choice between candidate best paths. The second table is also displayed in Appendix A.2.

Only 1 global candidate remains at the end: F♯ minor (3 sharps). Even if the piece is in A major instead of F♯ minor, the process found the correct global key signature, which is what impacts the pitch spelling. This information, along with the local tonalities computed in each measure and used in the second table to choose between paths with the same number of accidents, was critical in finally selecting a spelling. Indeed, for this piece, the algorithm reached an accuracy of 100% compared to the groundtruth [1].

3.4 Rewriting Passing Notes

After choosing a spelling with the algorithm of Section 3.3, we apply local corrections by rewriting the passing notes, using a slight generalisation of the rewrite rules proposed by D. Meredith in the original PS13 Pitch-Spelling algorithm [4], step 2.

Every rule applies to a trigram of notes ν_0, ν_1, ν_2, and rewrites the middle note ν_1, by changing its name. In Figure 6, we present the rules for particular cases of notes. In general however, the rules are defined by patterns comparing the respective note names of ν_0, ν_1, and ν_2 (without the accidentals), and the difference between their pitch (in number of semitones). For instance, in the left-hand-side C C♭ C of the first rule *broderie down*, ν_0, ν_1, and ν_2 all have the same note name C, the difference, in semitons, between ν_0 and ν_1 is -1 and the difference between ν_1 and ν_2 is $+1$. This rule rewrites the middle C♭ (ν_1) into B.

The rewrite rules are applied from left to right to the sequence spelled notes. Note that at each rewrite step, at most one rule can be applied.

3.5 Deterministic Variant

We propose a variant of the algorithm presented in Sections 3.2 and 3.3, which is more efficient but less exhaustive. This algorithm, called PS13b (as in "PS13 with bar info"), is very similar to PS13 [4], except that it uses the information on measures, which is assumed available in this paper but not in [4], in order to estimate global and local keys. In [4], that estimation is done (implicitely) by counting the number of occurrences of the (assumed) tonic note in a window whose optimal size was evaluated manually.

In the algorithm PS13b, the choice of the spelling with name n and accidental a for the input note ν_i (Section 3.2, transition rules (1)-(4)), is forced to the (unique) spelling in the chromatic harmonic scale of the current key K [1]. Hence, the transitions are deterministic, and there is no need to search for best spelling in a measure because there is only one. The rest of the algorithm works as described in Section 3.3. The complexity of the table construction in this case is $O(n \times p)$ where p is the total number of notes in input and n is the number of keys considered. This complexity is significantly better than the one of the exhaustive algorithm in Sections 3.2 and 3.3. In counterpart, some potentially correct spellings will be missed (see Section 4.3).

broderie down	C C♭ C	$\rightarrow$	C B C
broderie up	C C♯ C	$\rightarrow$	C D♭ C
descending_{11}	C C♭ A	$\rightarrow$	C B A
descending_{12}	C C♭♭ A♭	$\rightarrow$	C B♭ A♭
descending_{21}	C A♯ A	$\rightarrow$	C B♭ A
descending_{22}	C A♯ A♭	$\rightarrow$	C B♭ A♭
ascending_{11}	A A♯ C	$\rightarrow$	A B♭ C
ascending_{12}	A♭ A♯ C	$\rightarrow$	A♭ B♭ C
ascending_{21}	A C♭ C	$\rightarrow$	A B C
ascending_{22}	A C♭ C♯	$\rightarrow$	A B C♯

Figure 6. Rewrite rules for passing notes (particular cases).

4. EVALUATION

4.1 Implementation

The algorithms PSE (of Section 3.3) and PS13b (of Section 3.5) have been implemented [2] in C++20. This language was chosen for the sake of efficiency and for integration into larger systems. The implementation is object oriented, with general classes for pitches, keys, *etc*, and data structures specific to the algorithm, such as states, bags of best paths and tables.

The input must be provided by a note enumerator, associating to each natural number a midi pitch, a bar number, and a flag of simultaneity (with the next note). Therefore, our algorithm can be integrated in a larger project. This has been done for a MIDI-to-score transcription framework, where the timings (in particular the bar boundaries) are computed before pitch spelling.

A Python binding, based on pybind11 [16], was also written and used for evaluation. It offers calls (in Python) to the methods of the C++ implementation, for the various procedures and steps presented in Section 3.

For the evaluation, we used the Music21 toolkit [17], in association with the above Python binding. Music21

[1] See Appendices A.2 and A.3 for more details on the processing of Bach's Fugue BWV 864 (Example 3.3).

[2] The C++ code as well as the Python evaluation scripts are publicly accessible at https://gitlab.inria.fr/pse/pse.

parses the MusicXML files in the evaluation datasets (see Section 4.2) and extracts for each note the information needed by the algorithm:

- the MIDI key value,
- the number of the measure the note belongs to,
- a flag telling whether the note is simultaneous with the next one.

This information is fed to the PS procedure, and the spellings computed are compared to the ones in the original scores. Moreover, some scores are produced in output that highlight the errors of our procedure with colour codes, and mark each measure with the estimated local key [3].

4.2 Datasets

Two main datasets were used for performance evaluation: a monophonic (complex) one, originated from the Lamarque-Goudard rhythm textbook [18] *D'un Rythme à l'Autre*, and the ASAP piano dataset [19].

Performances of both algorithms described in Section 3 (PSE and PS13b) were assessed on the integrality of the Lamarque-Goudard dataset, containing 250 excerpts, as MusicXML files, from pieces of extremely various styles, from Bach and Scarlatti to Wolf, Duparc, Debussy, Ibert...

Evaluation was also executed on 5 separate corpora from the 222 pieces of the ASAP piano dataset, also in MusicXML format. All Bach preludes and fugues from the Well Tempered Clavier present in ASAP were used, except Preludes BWV 856 and 873 for technical reasons. All sonata movements by Mozart and Beethoven included in ASAP were also tested, as well as the K 475 Fantaisie by Mozart. Every one of the 13 Chopin Etudes contained in ASAP, from both opus 10 and 25, was used, as well as the 8 Rachmaninov preludes present, from both opus 23 and 32. The cumulated total of notes spelled by our two algorithms for this evaluation reaches a value of 216 464.

4.3 Results and Discussion

Experimentations were conducted for several combinations of the weight domains in $\{W_{\ell ex}, W_+\}$, with the best results obtained when using the weights of W_+. The execution time is about 1.68s on average per piece of the evaluation corpus (subset of Bach WTC), with the exhaustive algorithm PSE presented in Section 3.3, whereas it is only 0.04s on average per piece with the deterministic variant PS13b presented in Section 3.5, with results less accurate by more than 1%.

A summary of the evaluation results is presented in Table 1. The detailed results are also accessible online [3]. They are organised by corpus, each comprising a folder for each of the algorithm PSE and PS13b. Results for alternative versions of the algorithms corresponding to different ways of combining weights to compute costs (either lexicographically or additively) are also included). Every folder contains a table summarizing the results obtained on all the pieces in the considered corpus. In addition to the tables, folders relating to the Bach, Beethoven and Lamarque-Goudard corpora include the annotated MusicXML scores of the treated pieces, where the spelling errors are annotated with color codes, and green notes indicate an initial error corrected by the final rewriting. The annotations also include the global key estimation, and local key estimations for each measure. The score obtained after the execution of PSE on Bach's Fugue BWV 893 in b minor is included in Appendix A.3 as an example.

Regarding global and local tonality estimation, our algorithm achieves very good results when we compare key signatures together but tends to prefer minor tones to their major relatives. This can be explained by the large number of notes a minor tonality possesses in our acceptation, as we accept spelling from both harmonic and natural minor modes, as well as the ascendant melodic one. The only error of global tone estimation on our whole Well-Tempered Clavier dataset is directly due to this tendency: the presence of natural B's in the BWV 870 prelude (C major of book 2), in a piece where flat Bs are also numerous due to modulations to F major, D minor *etc.*, did not prevent our algorithm from estimating the piece as written in D minor, because these natural B's were interpreted as part of the ascendant melodic minor mode of D, instead of indicators of a C major context.

About enharmonic tones, which are absolutely impossible to distinguish when only pitches and durations of notes are given, if the algorithm only proposed the correct global tone or its enharmonic counterpart among its global candidates at the end of the first pass, and if its final estimated global key is enharmonic to the correct one, then the piece is renamed in the enharmonic rival global tone and errors are computed on this version. This way of treating that issue is in accordance with the definition of a well spelled piece by [6], also shared with [4], and [2], relying on correctness of the intervals of the piece.

4.4 Comparison With Other Systems

On the task of global tonality guessing (KE), we have compared our algorithms' performances to the ones obtained with the Krumhansl-Schmuckler (KS) model for key determination, as implemented in the Music21 Python library [17]. This famous key-finding algorithm computes for every major and minor tonality a correlation coefficient between profile values of the tested key and total durations of their corresponding pitch class in the musical piece considered. It then chooses the best tonality according to the calculated correlation coefficients. It is interesting to note that our algorithms only need to know measure delimitations and not note durations whereas KS uses note durations and does not care about measures. On the whole corpus (both ASAP and Lamarque-Goudard) we attain a 93% correctness of key signature determination on average, while KS obtains a 75% accuracy in total.

The results for global tonality guessing (KE) on the Lamarque-Goudard (LG) dataset are is rather low, in comparison to ASAP. The LG corpus, extracted from a rhythm textbook, consists of 250 short excerpts of longer pieces.

[3] See https://github.com/florento/PSEval/ for the evaluation results, including annotated scores and tabular summaries.

Table 1. Accuracy results of both pitch spelling algorithms on pieces from widely different styles and correctness of their global key signature estimation, compared to Krumhansl-Schmuckler algorithm's performances (Music 21 implementation).

	number of spelled notes	pitch spelling PSE	pitch spelling PS13b	key estimation PSE	key estimation PS13b	key estimation Krumhansl-Schmuckler
Bach WTC ASAP corpus	55530	99.50%	98.27%	99.09%	98.29%	87.27%
5 movements from Mozart Sonatas present in ASAP	10043	99.11%	97.30%	100%	100%	80%
Fantaisie K. 745 plus 5 movements from Mozart Sonatas	13830	97.65%	95.97%	80%	80%	60%
33 movements from Beethoven Sonatas	87292	97.64%	95.65%	92.32%	95.71%	66.15%
13 Etudes by Chopin	25103	96.71%	96.03 %	96.15%	96.15 %	84.62%
4 Rachmaninov Preludes	7022	98.76%	97.49%	100%	100%	100%
Lamarque-Goudard	27687	98.46%	98.23%	76.90%	74.30%	50.60%

The pitches and duration of notes appearing in extracts are not necessarily representative of the whole pieces, which explains why KS performs rather poorly on it, since its correlation coefficients with tonal profiles become erroneous. Our algorithms, mainly relying on accidentals number minimization to infer the tonality, therefore prove significantly more robust when it comes to shorter extracts.

We do not provide a comparison table on Pitch Spelling results for several reasons. First, we assume given measure information, unlike the algorithms PS13 [4], CIV [8] or PKSpell [9]; to this respect, a comparison would not be fair. Second, most of the former evaluations of pitch spelling algorithms used as a benchmark the Musedata dataset proposed by D. Meredith [4]. We could not evaluate our algorithms on this dataset because it does not include the measure boundary information required by our procedures. Since note durations are included in Musedata, it would be possible to conduct an evaluation on this dataset by manually providing a time signature for each of its 216 pieces. Nevertheless, with success rates (for PS) of 99.41% with PS13 [4], 99.82% with CIV [8] and 99.87% with PKSpell [9], the remaining room for improvements on this benchmark is rather marginal, and we preferred to focus on larger datasets like ASAP to evaluate and improve our algorithms.

The recent system PKSpell [9], for which it is reported a 0.13% error rate on MuseData (the best results so far), has also been evaluated on 33 pieces of the challenging piano dataset ASAP [19]. It shows on this dataset an accuracy of 96.50% for the pitch spelling task and 90.30% for key signature estimation. Since the identities of the 33 pieces for evaluation are not disclosed in [9], it is not feasible for this paper to report performances on the exact same pieces. However, with an accuracy of 98.19% on average for pitch spelling on 110 pieces (by 5 different composers ranging from Bach to Rachmaninov) from the same ASAP dataset, as reported on Table 1, and 95.58% for global key signature estimation, it is likely that the proposed PSE algorithm (and PS13b) should at least have similar performances to PKSpell, if not better.

5. CONCLUSION

We have presented two algorithms, PSE and PS13b for joint pitch spelling and estimation of global and local keys, from MIDI data including information on measure boundaries. Originally thought to be integrated in a transcription framework, these procedures could also be used in various tasks of music notation processing. Since PS13b has proven to be very efficient, it could also be used for displaying music notation from MIDI data in real-time.

The evaluation on challenging datasets has shown robust results both for pitch spelling and key signature estimation. Regarding the estimation of keys, there is currently a bias towards minor tonalities which are often preferred to their major relative tones that should be corrected. However it does not impact the chosen key signature.

Several directions can be explored in order to improve the current approach, such as a refinement of the weight domain for taking into account note durations and metric weight (strong or weak beats) when computing the best path in a given measure. Moreover, in order to improve the accuracy of the tonal analysis, some subtler musical criteria could be implemented, such as cadence detection or chord classification as well as a process to detect justified key signature changes.

As outlined in Section 2.2, the only parameters related to durations that we are considering in our algorithms are the division into measures and knowledge about the simultaneity of notes (which in our case only depends on their onsets). The results of our experiments are already solid, although the seemingly important parameter of *individual note durations* is ignored. We shall try to include this parameter to see how it can include the results, however, it is not clear to us what weighting to give to durations in the search of best spellings presented in Section 3.2, while keeping our model independent of the corpora considered.

Another way to extend our approach to other genres, perhaps less tonal, would be to integrate new *modes* into the computation of the PS table. For instance, one may consider the integration of jazz modes (*e.g.*, Ionian, Dorian *etc*) in order to tackle the problem of pitch spelling for jazz, which has not been studied at lot in the literature. This could be of interest in particular for the notation of jazz soli, improvisations, and bass lines for instance.

6. REFERENCES

[1] J. Nagel, "The chromatic modal scale: Proper spelling for tonal voice-leading," *JOMAR Press*, 2007.

[2] D. Temperley, *The cognition of basic musical structures*. MIT press, 2004.

[3] E. Chew and Y.-C. Chen, "Real-time pitch spelling using the spiral array," *Computer Music Journal*, vol. 29, no. 2, pp. 61–76, 2005.

[4] D. Meredith, "The PS13 pitch spelling algorithm," *Journal of New Music Research*, vol. 35, no. 2, pp. 121–159, 2006.

[5] E. Cambouropoulos, "Pitch spelling: A computational model," *Music Perception*, vol. 20, no. 4, pp. 411–429, 2003.

[6] A. K. Honingh, "Compactness in the Euler-lattice: A parsimonious pitch spelling model," *Musicae Scientiae*, vol. 13, no. 1, pp. 117–138, 2009.

[7] B. Wetherfield, "The minimum cut pitch spelling algorithm: Simplifications and developments," in *TENOR*, 2020.

[8] G. Teodoru and C. Raphael, "Pitch spelling with conditionally independent voices," in *Proc. conf. of the Int. Society for Music Information Retrieval (ISMIR)*, 2007, pp. 201–206.

[9] F. Foscarin, N. Audebert, and R. Fournier-S'Niehotta, "PKSpell: Data-driven pitch spelling and key signature estimation," in *Proc. 22nd conf. of the Int. Society for Music Information Retrieval (ISMIR)*, 2021.

[10] *Official MIDI Specifications*, MIDI Association, 2023. [Online]. Available: https://www.midi.org/specifications/midi-2-0-specifications

[11] G. Weber, *Versucht einer geordneten Theory der Tonsetzkunst*. B. Schott's Söhne, 1818.

[12] L. Feisthauer, L. Bigo, M. Giraud, and F. Levé, "Estimating keys and modulations in musical pieces," in *Sound and Music Computing Conference (SMC)*, 2020.

[13] E. Gould, *Behind Bars: The definitive guide to music notation*. Faber Music, 2011.

[14] L. Huang, "Advanced dynamic programming in semiring and hypergraph frameworks," in *COLING*, 2008.

[15] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[16] J. Wenzel. pybind11 – Seamless operability between C++11 and Python. [Online]. Available: https://pybind11.readthedocs.io

[17] M. S. Cuthbert and C. Ariza, "music21: A toolkit for computer-aided musicology and symbolic music data," in *Proc. conf. of the Int. Society for Music Information Retrieval (ISMIR)*, 2010.

[18] E. Lamarque and M.-J. Goudard, *D'un Rythme à l'autre*. Lemoine, 1997, vol. 1-4.

[19] F. Foscarin, A. Mcleod, P. Rigaux, F. Jacquemard, and M. Sakai, "ASAP: A dataset of aligned scores and performances for piano transcription," in *Proc. conf. of the Int. Society for Music Information Retrieval (ISMIR)*, 2020, pp. 534–541.

A. APPENDIX

In this appendix, we present some details about the procedure (Section A.1) and two samples of evaluation results, for illustration. The complete evaluation results may be found at https://github.com/florento/PSEval/.

A.1 Table of Weber

The table of relationship of keys defined in [11], and used in Section 3.3, step 3, is displayed below. Keys in major mode are uppercase, keys in minor mode are lowercase.

KS		-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
		C♭	G♭	D♭	A♭	E♭	B♭	F	C	G	D	A	E	B	F♯	C♯	a♭	e♭	b♭	f	c	g	d	a	e	b	f♯	c♯	g♯	d♯	a♯
-7	C♭	0	1	2	2	3	4	4	5	6	6	7	8	8	9	10	1	2	3	3	4	5	5	6	7	7	8	9	9	10	11
-6	G♭	1	0	1	2	2	3	4	4	5	6	6	7	8	8	9	2	1	2	3	3	4	5	5	6	7	7	8	9	9	10
-5	D♭	2	1	0	1	2	2	3	4	4	5	6	6	7	8	8	2	2	1	2	3	3	4	5	5	6	7	7	8	9	9
-4	A♭	2	2	1	0	1	2	2	3	4	4	5	6	6	7	8	1	2	2	1	2	3	3	4	5	5	6	7	7	8	9
-3	E♭	3	2	2	1	0	1	2	2	3	4	4	5	6	6	7	2	1	2	2	1	2	3	3	4	5	5	6	7	7	8
-2	B♭	4	3	2	2	1	0	1	2	2	3	4	4	5	6	6	3	2	1	2	2	1	2	3	3	4	5	5	6	7	7
-1	F	4	4	3	2	2	1	0	1	2	2	3	4	4	5	6	3	3	2	1	2	2	1	2	3	3	4	5	5	6	7
0	C	5	4	4	3	2	2	1	0	1	2	2	3	4	4	5	4	3	3	2	1	2	2	1	2	3	3	4	5	5	6
1	G	6	5	4	4	3	2	2	1	0	1	2	2	3	4	4	5	4	3	3	2	1	2	2	1	2	3	3	4	5	5
2	D	6	6	5	4	4	3	2	2	1	0	1	2	2	3	4	5	5	4	3	3	2	1	2	2	1	2	3	3	4	5
3	A	7	6	6	5	4	4	3	2	2	1	0	1	2	2	3	6	5	5	4	3	3	2	1	2	2	1	2	3	3	4
4	E	8	7	6	6	5	4	4	3	2	2	1	0	1	2	2	7	6	5	5	4	3	3	2	1	2	2	1	2	3	3
5	B	8	8	7	6	6	5	4	4	3	2	2	1	0	1	2	7	7	6	5	5	4	3	3	2	1	2	2	1	2	3
6	F♯	9	8	8	7	6	6	5	4	4	3	2	2	1	0	1	8	7	7	6	5	5	4	3	3	2	1	2	2	1	2
7	C♯	10	9	8	8	7	6	6	5	4	4	3	2	2	1	0	9	8	7	7	6	5	5	4	3	3	2	1	2	2	1
-7	a♭	1	2	2	1	2	3	3	4	5	5	6	7	7	8	9	0	1	2	2	3	4	4	5	6	6	7	8	8	9	10
-6	e♭	2	1	2	2	1	2	3	3	4	5	5	6	7	7	8	1	0	1	2	2	3	4	4	5	6	6	7	8	8	9
-5	b	3	2	1	2	2	1	2	3	3	4	5	5	6	7	7	2	1	0	1	2	2	3	4	4	5	6	6	7	8	8
-4	f	3	3	2	1	2	2	1	2	3	3	4	5	5	6	7	2	2	1	0	1	2	2	3	4	4	5	6	6	7	8
-3	c	4	3	3	2	1	2	2	1	2	3	3	4	5	5	6	3	2	2	1	0	1	2	2	3	4	4	5	6	6	7
-2	g	5	4	3	3	2	1	2	2	1	2	3	3	4	5	5	4	3	2	2	1	0	1	2	2	3	4	4	5	6	6
-1	d	5	5	4	3	3	2	1	2	2	1	2	3	3	4	5	4	4	3	2	2	1	0	1	2	2	3	4	4	5	6
0	a	6	5	5	4	3	3	2	1	2	2	1	2	3	3	4	5	4	4	3	2	2	1	0	1	2	2	3	4	4	5
1	e	7	6	5	5	4	3	3	2	1	2	2	1	2	3	3	6	5	4	4	3	2	2	1	0	1	2	2	3	4	4
2	b	7	7	6	5	5	4	3	3	2	1	2	2	1	2	3	6	6	5	4	4	3	2	2	1	0	1	2	2	3	4
3	f♯	8	7	7	6	5	5	4	3	3	2	1	2	2	1	2	7	6	6	5	4	4	3	2	2	1	0	1	2	2	3
4	c♯	9	8	7	7	6	5	5	4	3	3	2	1	2	2	1	8	7	6	6	5	4	4	3	2	2	1	0	1	2	2
5	g♯	9	9	8	7	7	6	5	5	4	3	3	2	1	2	2	8	8	7	6	6	5	4	4	3	2	2	1	0	1	2
6	d♯	10	9	9	8	7	7	6	5	5	4	3	3	2	1	2	9	8	8	7	6	6	5	4	4	3	2	2	1	0	1
7	a♯	11	10	9	9	8	7	7	6	5	5	4	3	3	2	1	10	9	8	8	7	6	6	5	4	4	3	2	2	1	0

A.2 Execution Tables for the Fugue BWV 864

We display below the verbatim of the tables computed when processing the Fugue BWV 864, see Example 3.3.

First table computed for processing the Fugue BWV 864 (Example 3.3):

```
spelling 813 notes
PSE: first table
Row Costs:
Gbmajor (6b) cost accid=199 dist=0 chromarm=188 color=10  cflat=25
Dbmajor (5b) cost accid=251 dist=0 chromarm=237 color=19  cflat=46
Abmajor (4b) cost accid=291 dist=0 chromarm=280 color=35  cflat=44
Ebmajor (3b) cost accid=279 dist=0 chromarm=274 color=123 cflat=24
Bbmajor (2b) cost accid=252 dist=0 chromarm=258 color=246 cflat=8
Fmajor  (1b) cost accid=219 dist=0 chromarm=222 color=274 cflat=0
Cmajor  (0)  cost accid=169 dist=0 chromarm=175 color=344 cflat=0
Gmajor  (1#) cost accid=123 dist=0 chromarm=130 color=3   cflat=4
Dmajor  (2#) cost accid=71  dist=0 chromarm=78  color=3   cflat=4
Amajor  (3#) cost accid=38  dist=0 chromarm=43  color=2   cflat=4
Emajor  (4#) cost accid=69  dist=0 chromarm=81  color=1   cflat=4
Bmajor  (5#) cost accid=113 dist=0 chromarm=124 color=1   cflat=4
F#major (6#) cost accid=154 dist=0 chromarm=165 color=0   cflat=0
Ebminor (6b) cost accid=179 dist=0 chromarm=205 color=23  cflat=19
Bbminor (5b) cost accid=220 dist=0 chromarm=257 color=30  cflat=34
Fminor  (4b) cost accid=255 dist=0 chromarm=283 color=76  cflat=29
Cminor  (3b) cost accid=229 dist=0 chromarm=279 color=159 cflat=12
Gminor  (2b) cost accid=204 dist=0 chromarm=262 color=293 cflat=4
Dminor  (1b) cost accid=167 dist=0 chromarm=222 color=293 cflat=0
Aminor  (0)  cost accid=130 dist=0 chromarm=175 color=344 cflat=0
Eminor  (1#) cost accid=110 dist=0 chromarm=130 color=3   cflat=4
Bminor  (2#) cost accid=67  dist=0 chromarm=78  color=3   cflat=4
F#minor (3#) cost accid=33  dist=0 chromarm=43  color=2   cflat=0
C#minor (4#) cost accid=69  dist=0 chromarm=81  color=1   cflat=4
G#minor (5#) cost accid=111 dist=0 chromarm=124 color=1   cflat=5
D#minor (6#) cost accid=145 dist=0 chromarm=165 color=0   cflat=0
```

Second table computed for processing the Fugue BWV 864 (Example 3.3):

```
Row Costs:
Amajor  (3 sharps) cost accid=38 dist=35 chromarm=3 color=4 cflat=1
F#minor (3 sharps) cost accid=33 dist=21 chromarm=0 color=0 cflat=0
```

A.3 Annotated Score of the Fugue BWV 893

The Fugue BWV 864, in the previous section and Example 3.3, was spelled with PSE with an accuracy 100%, see

https://github.com/florento/PSEval/blob/941d7e10731fefdcd28cdac1c53a58cd28d132f2/Results_ASAP/Bach/PSE/BWV864_Fugue.musicxml

Each time that one our algorithms makes some spelling errors, our evaluation script produce a copy of the original (musicXML) score with the mistakes annotated. For instance, the four errors in the spelling obtained with PSE for the Fugue BWV 893 from the ASAP dataset are highlighted in red in the following score:

https://github.com/florento/PSEval/blob/941d7e10731fefdcd28cdac1c53a58cd28d132f2/Results_ASAP/Bach/PSE/BWV893_Fugue.musicxml

or, for a PDF version:

https://github.com/florento/PSEval/blob/941d7e10731fefdcd28cdac1c53a58cd28d132f2/Results_ASAP/Bach/PSE/BWV893_Fugue.pdf

All scores processed with our method for the evaluation presented in Section 4, for the two datasets of Section 4.2 (see Table 1 for the list), with annotations as well as tables of results, are available at

https://github.com/florento/PSEval/

MORTON FELDMAN'S "PROJECTIONS ONE TO FIVE" – EXPLORING A CLASSICAL AVANT-GARDE NOTATION BY MATHEMATICAL REMODELLING

Markus Lepper
semantics gGmbH, Berlin, DE
post@markuslepper.eu

Baltasar Trancón y Widemann
semantics gGmbH, Berlin, DE
Technische Hochschule Brandenburg a.d. Havel, DE

ABSTRACT

The compositions *Projection 1* to *Projection 5* by Morton Feldman are an important milestone in the application of graphical notation. The meta language `tscore` allows easy construction of a computer model of the original scores. On this model, automated performance, graphical rendering, and different analyses can be applied. The practical implementation work brings up the peculiarities of the original notational meta-model and scores, which without this effort are easily overlooked.

1. THE ORIGINAL SCORES

The compositions *Projection 1* to *Projection 5* were composed by Morton Feldman in 1950/51 and published by Edition Peters from 1959 to 1964. The format is that of a graphical score which only specifies the time position, duration, and the pitch register of the events, see Figure 1. The selection of the sounding pitches is left to the players. These works are among the earliest compositions with such "indeterminate pitches". "Feldman's graphic scores of the early 1950s are important [. . .] as works whose wide influence extended over the next few decades." [2, p. 10]

The project presented in this article takes a closer look to the syntax and the possible semantics of the "Projections". By constructing mathematical meta-models and realizing them as software, properties, prerequisites and problems become visible which can easily be overlooked otherwise, and indeed often have been. "[These works] have a certain iconic status in the popular and scholarly literatures, although very few have been analysed in close detail."[2, p. 10] Not before 2016 such more detailed analyses appeared [3].

Each piece is for a different selection of instruments, [1] but all share the same newly invented graphical score format, the intended sound character ("Dynamics are very low", except in no. 1), and even the tempo (BPM $\approx$ 72).

[1] Determined by practical considerations, see [3, p. 18, 20]. Vigil states that each piece presents even a different *type* of ensemble. [1, p. 234] Feldman regarding the title: "My desire here was not to 'compose,' but to project sounds into time." ('Liner notes' in [4]) Boutwell sees a connection to Varèse [5, p. 465, 477]; Further considerations in [3, p. 24].

Despite its novel graphical appearance, the format indeed adheres widely to the conventions of standard notation, as voices appear vertically stacked and time flows from left to right. The staff for each instrument is a free space, into which non-overlapping rectangles are drawn, representing one played event each. According to the explanatory forewords (see below), the onsets and the offsets of these shall be read as aligned to quarters (= quarter notes = quarters of a box width), which are not explicitly represented visually.

The height of the event symbols is placed in the lower, middle, or upper third of the free space of the staff, indicating the duration for which an arbitrary pitch from the lower, middle, or upper sound register may be played.

So the base of the score is a fixed *grid*: Each measure has the same musical length (four pulses), and the same graphical width, and each page (beside possibly the last) holds the same number of measures, (Exception is no. 4 which changes between 10 and 11, see below.)

The measures are limited by vertical measure bars, which extend from the highest to the lowest staff which carries events in an adjacent measure. These bars are dotted lines and make up the vertical limits of the "boxes", mentioned as such in the explanatory forewords, but correspond to a conventional "measure". [2]

The events for the strings are further qualified as ordinary, harmonic, pizzicato, or sul ponticello. In nos. 1 and 4 only the first three appear and are distributed to three different staves. In no. 4 the solo violin has (in all three voices arco, pizzicato, and harmonics) few multi-stops. These are attributes with the number of pitches go be played, as are events in all piano voices. In all pieces, one of two piano staves is for audible key presses, an additional staff is for silently pressed harmonics. These shall resonate not only to the piano sounds but also to other ensemble instruments—the afterword to no. 2 says "Trumpet plays into open piano".

The staff for the audible piano events is the only one in which events may overlap horizontally, what happens seldomly (no. 2 ms. 56, 63, 65, 82; no. 3 ms. 7, 25; no. 4

[2] These measures are for orientation only and do in no way imply metric emphasis or structure. [3, p. 92] In the open and extensible "LMN catalogue" of basic notation properties this can be described as NOTA.QUANDO.SEPARATORMETRIMENDAX or TEMPUS.COLORATUM.ABUSU. Currently covering only conventionally used Common Western Notation, LMN ([6], [7]) nevertheless lists about 890 different properties, defined by mathematical meta-models. These range from trivial facts to complex algorithms and are referred to by hierarchically structured identifiers, in analogy to zoology nomenclature etc. These are built on terms from the Latin language to restrict them to their role as mere labels for mathematical definitions and to avoid misleading connotations.

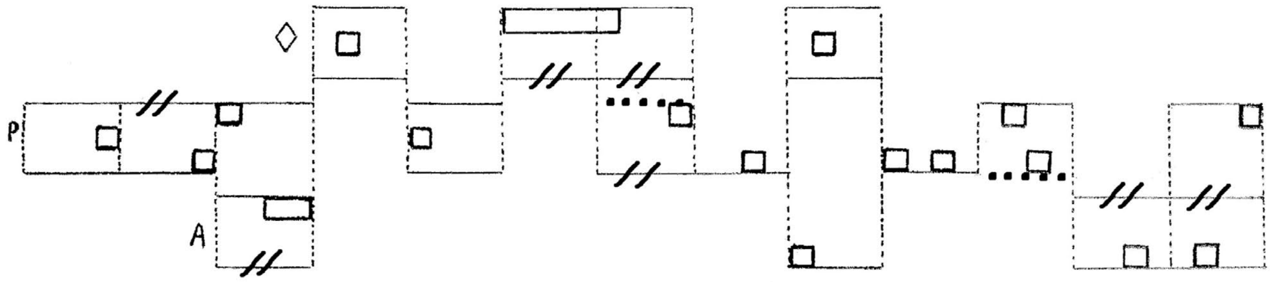


Figure 1. The first original page of Feldman, *Projection 1,* as printed by Peters, plus our correction marks "// " and "• • •".

ms. 25, 29. 33, 38, 44, 63, 64, 67, 77, 82; no. 5 ms. 2, 5, 17, 25, 34) and never with more than two sounds. [3]

In the work of Feldman, the usage of this style of graphic notation was only an intervening period, but an important one: It opened new ways of composing beyond the traditional focus on pitch classes. When the composer returned to exact notation, he had learned a much more free approach to pitch composition. "Feldman's conventionally notated music is himself playing his graph music", said John Cage, according to [8, p. 216].

2. ISSUES WITH THE ORIGINAL SCORES

The scores of all five works, as published by Peters separately, have some remarkable issues:

(Explicit definitions) Due to the radically new nature of the applied notation, this is defined explicitly by a short *explanation*, not more than one page, except for no. 4. So we have the rare and rewarding case that the semantics of notational devices are explained explicitly. [4] This page appears as prefix or postfix, is integrated into the page numbering or not, is written by hand or by type-writer, see Table 1.

(Edition errors) Proof-reading has been rather sloppy: The role of the figures with the piano notation (namely to give the number of keys to press simultaneously) is explained in nos. 3 to 5, but not in no. 2.

No. 2 is printed without the date of publication.

The title page of no. 5 says "for 3 Flutes, Trumpet, Piano and 3 Violoncelli", but indeed the score contains *two* pianos.

The staves in no. 5 are labelled with the instruments' names, except for the very first page. [5]

(Turning points) The layout of the scores is somewhat careless: Empty pages are interspersed in a way not to minimize the turning points but to maximize them.

(Empty bars) The score seems to be written by hand on a *transparent paper*, with a grid paper underlying and ruling the writing—in those days a common technique in architecture and engineering. But taking the grid away, which gives the look as it is printed, implies a *fundamental shift of paradigms*: Two adjacent empty measures (see no. 1, page 4, measures 48 and 49), and any empty measure adjacent to a line break, are not longer recognizable in a pure syntactic way! While everywhere else the sequence of measures is unambiguously recognizable even when written on rubber and arbitrarily deformed, [6] now the distances as such carry semantics.

Notably, the hand-written date of completion appearing at the end of all scores (except no. 5) does indeed *carry musical semantics*, because it indicates that no empty measures follow! [7]

(Horizontal lines) The horizontal separation lines get a very different treatment through the five pieces. This is discussed in detail in section 5.

3. REAL-WORLD PERFORMANCE ISSUES

With the piano: There are measures in which the pianist needs three hands! By automated data analysis (see section 7) we found that this is the case only in no. 2, measures 56, 63, 65, and 82. All these situations imply that there is one chord pressed in the "harmonic" piano staff, overlapping with two chords in the "sounding" staves. The occurring patterns are

```
  56               63/65            82
 YYYYY              YYYY           XXXX
XXXXXXX           XXXX             YYYY
 HHHHHHHH      HHHHHHHHHHH     HHHHHHHH
```
[= cases e)/g)   b)/g)   e)/g) from Table 2]

All patterns begin with some solo chord (sound or harmonics), so they can easily be realized with a *sostenuto pedal* (= "Steinway pedal"): This very first chord is taken into this pedal and the hand is free again.

If such a pedal is not available, the normal *sustain pedal* must be used instead. Of course this may only be active for the minimal possible interval: Short before the third chord must be played, that one of the sounding chords which has the earlier end ("XXXX" above) is taken into the pedal, which will be released with its end.

Table 2 shows all combinations for both kinds of pedals, especially the required silent returns. If the second and third chord come synchronously (like in ms. 82 and cases e) and f) in the Table) the first chord must be taken into the pedal anyhow. If it lasts longer than at least one of

[3] Found by automated analysis, see section 7. Discussed also in [3, p. 120p.].
[4] LMN property INFRA.REGULA.DICTA [6, p. 522].
[5] See also [1, Endnote 3].
[6] In the LMN catalogue this is the property NOTA.NONQUANTIBUS [6, p. 96].
[7] Cline discusses the role of an empty "box" at the end of the related work "Intersection I" (1951) [3, p. 176].

number	instruments	kinds of instruments	players	position of explanation	page number of expl.	medium of expl.	inter-staff gap	horizontal delimiter	measures per page	pages with measures	measures on last pg.	total measures	total duration at BPM 72	first staves labeled	subsequent staves labeled
1	Vcl	1	1	pre	–	hand	1	free	13	4	12	51	2’50	yes	no
2	Fl, Trp, Vl, Vcl, Pft	5	5	post	10	typewriter	0	free	10	9	5	85	4’43	yes	idem
3	2 Pft	1	2	post	–	typewriter	0	free	10	3	8	28	1’33	yes	idem
4	Vl, Pft	2	2	pre	I-II	hand	0.5; 2.5	reg	10; 11	8	–	84	4’40	yes	no
5	3 Fl, Trp, 3 Vcl, 2 Pft	4	9	pre	–	typewriter	0	solid	10	4	–	40	2’13	no[sic!]	yes

Table 1. Properties of *Projection 1* to *Projection 5*

$$t_1 \leq t_2 \leq t_3 < u_1, u_2, u_3$$

a)	$t_1 \leq t_2 < t_3$	$h_1 \neq h_2$	$p = t_3 - \epsilon$	$u_1 < u_2 \wedge u_1 \leq u_3$	$h_3 = h_1$	$p^* = u_1$	
b)	$t_1 \leq t_2 < t_3$	$h_1 \neq h_2$	$p = t_3 - \epsilon$	$u_2 < u_1 \wedge u_2 \leq u_3$	$h_3 = h_2$	$p^* = u_2$	
c)	$t_1 \leq t_2 < t_3$	$h_1 \neq h_2$	$p = t_3 - \epsilon$	$u_1 = u_2 \wedge u_1 \leq u_3$	$h_3 = h_x$	$p^* = u_2$	
d)	$t_1 \leq t_2 < t_3$	$h_1 \neq h_2$	$p = t_3 - \epsilon$	$u_3 < u_1 \wedge u_3 < u_2$	$h_3 = h_x$	$p^* = u_3$	$R_{x,3} = p^* - \epsilon$
e)	$t_1 < t_2 = t_3$	$h_2 \neq h_3$	$p = t_2 - \epsilon$	$u_1 \leq u_2 \wedge u_1 \leq u_3$		$p^* = u_1$	
f)	$t_1 < t_2 = t_3$	$h_2 \neq h_3$	$p = t_2 - \epsilon$	$u_y < u_1$		$p^* = u_y$	$R_{1,y} = p^* - \epsilon$
g)	$t_1 < t_2 \leq t_3$		$t_1 < q < t_2$			$q^* = u_1$	
h)	$t_1 = t_2 < t_3$		$t_1 < q < t_3$	$u_1 < u_2$	$h_3 = h_1$	$q^* = u_1$	
i)	$t_1 = t_2 < t_3$		$t_1 < q < t_3$	$u_2 < u_1$	$h_3 = h_2$	$q^* = u_2$	
j)	$t_1 = t_2 < t_3$		$t_1 < q < t_3$	$u_1 = u_2$	$h_3 = h_x$	$q^* = u_x$	

Explanation:
t_n = the start timepoints of a chord, given by the index
h_n = the hand used to play the chord with the same index
u_n = the end timepoint of the chord with that index
p = the timepoint of pressing down the (normal/sustain) pedal
p^* = the timepoint of releasing the (normal/sustain) pedal
q = the timepoint of pressing down the Steinway pedal
q^* = the timepoint of releasing the Steinway pedal
$R_{a,b}$ = the timepoint of returning silently to the chord number a, letting go chord number b
$t - \epsilon$ = right before t
$x \in \{1, 2\}$ $y \in \{2, 3\}$

Table 2. The different cases when playing three chords with only two hands.

the others (=case f), a hand must re-press its keys silently, immediately before the pedal is released, see Figure 2. A similar method must be applied when the third chord is shorter than all others (case d). These cases do not occur in the composition, but are likely when composing in this style.

The use of the normal pedal blurs the resonance effects intended with the “harmonics” voice. Such situations occur only in emphProjection 2, which is the first in the cycle which uses a piano. Apparently by the time the composer became aware of this effect and avoided it. [8]

With the strings: Programmed analysis shows that multi stops are only applied in the violin in no. 4. In the “harmonic” voice one double stop in ms. 32, in the “arco” voice on triple stop in ms. 17, both in the middle register. The “pizzicato” voice contains the most, which is appropriate to the nature of the tone production, namely two double, three triple and even one quadruple stop. Per register:

	2	3	4
hi		1	1
mid	1		
lo	1	2	

Problematic are the two triple stops in the low register: The lowest solution are the open string G3+D4+A4 (in MIDI notation), the top of which comes very close to any possible “mid” register.

The opposite problem with the quadruple stop in the high register: all four strings must be involved, over a distance of nearly two octaves of their tunings, which requires a wide definition of that register to allow the left hand to compress all pitches into it.

Comprehension: It appears as if the new way of composing was practised by Feldman in a primely more ab-

[8] In the subsequent works “Intersection 2 and 3” the piano part became again much more virtuoso, see [3, p. 40].

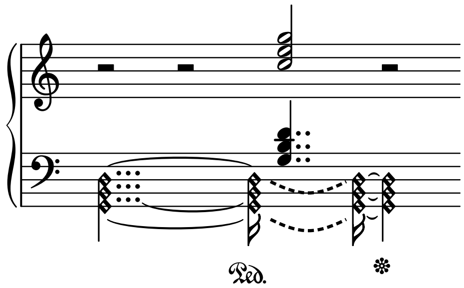


Figure 2. A method for playing overlapping piano events according to case f) in Table 2.

stract and programmatical way, not treating very thoroughly the issues of a concrete realization. [9] Since the idea of letting the pitches to the decision of the players was utmost revolutionary anyhow and dominated all considerations, this is understandable.

4. THE TSCORE MODEL

`Tscore` is a meta-meta-model for constructing models for the syntax and semantics of musical artefacts. [11] [12] [13] Figure 3 shows its simple fundamental paradigm in the style of entity–relationship. It may be surprising how well-known components of a model naturally exist on *different layers* of the architecture: The fundamental formula for representation (last line in the figure) is the core of the meta-meta-model. It states that every event • is identified by a pair of values: a voice from V and a timepoint from T. Each event relates to a collection of values from the parameter domains $D_a, D_b, \ldots$, indexed by parameter names from P. The events and the identity of the voices V change with every model, but all other types are defined with the meta-model—in `tscore` even the time structure is pluggable.

To encode and process all composition *Projection 1* to *Projection 5*', only one single meta-model had to be developed, the instances of which are the models of these pieces. Table 3 shows the "timeless" parameters used to adapt the meta-model to the needs of the different pieces. [10]

The complete code for parsing and evaluating are 275 lines of code (measured with "cloc" [14], not including the `main(..)` wrapper and file reading). plus about 400 loc for the graphical and 300 lines for the acoustic rendering. Plus arbitrary many for wanted analyses. All code (except the score data itself) is in the public domain; a publication as open source is under preparation.

Figure 4 shows the beginning of the `tscore` source text of no. 2: The events are realized by the codes "`h`", "`m`", and "`l`", for the high, middle, or low register. The events in the voices "`vl`" and "`vc`" for the string instruments may carry an additional character as qualifier: "`h`" for harmonics, "`p`" for pizzicato, and "`P`", for sul ponticello. In nos. 1 and 4, only the first two are used—realized in the Peters score not

[9] See also [9, p. 73] and the detailed discussion w.r.t "Untitled Composition" in [10, p. 33].

[10] The names have been chosen from the Latin language for smooth integration into the LMN catalog.

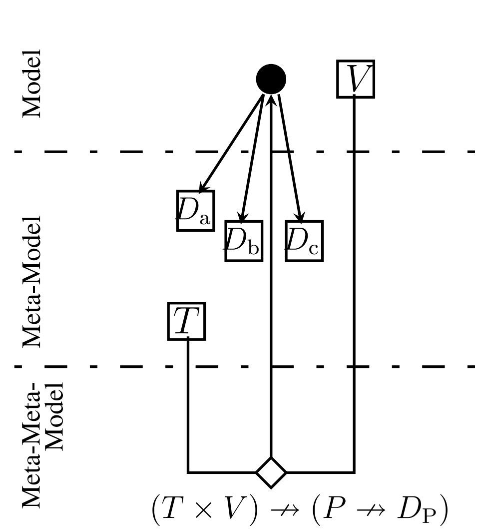


Figure 3. The fundamental paradigm of `tscore`: an event-based co-algebraic data model

by modifiers but by distributing the events to three separate systems. In no. 4 the solo violin may carry an additional digit to indicate multiple stops. The modifier "A" for arco/ordinario from the original score is realized in the `tscore` version as the default case = entered with no visible modifier. The piano voices must carry an additional decimal number of the keys to press.

The format of each voice is declared to the `tscore` parser by the timeless property `addenda`.

The rhythm is defined by (a) the division of the measure into 1, 2, or 4 entries, supported by (b) positive or negative dotting notation (see measures 3 to 6 of voice `fl`), plus (c) the rule that every event lasts up to the onset of its successor in the same voice. This may by a pause "`%`" or a prolongation "`-`". It took only one hour to enter the complete no. 2, including proofreading supported by the graphic output.

Since `tscore` defines "events" by a unique combination of "voice" and "timepoint", the piano staffs which contain timely overlapping events have been modelled by multiple voices (in fact: only two). Both voices are entered and processed separately, but written into the same staff / the same synthesizer channel when rendering. This is controlled by the timeless voice parameter `cumLinea`, see Table 3.

5. GRAPHICAL RENDERING OF THE MODELS

Our interest in the notation of Feldman's grid music was inspired by the Neoscore project, a new graphical music notation library [15] which used *Projection 2* as a test case. The idea suggested itself to employ a `tscore` model as the front-end, for more comfortably generating the data. Indeed the co-algebraic and visitor-based architecture of `tscore` allowed to implement the export into the Python data readable by Neoscore by just 40 lines of Java code—see Figure 5 for an example output of this co-operation.

The Neoscore rendering is only a proof-of-concept prototype, specially addressing no. 2. But by using `tscore`

timeless parameters per score:
• `visuum.interVoces` : $\mathbb{Q}$
// = gap between staves, given as a factor of event height
• `visuum.tactiInPagina` : $\mathbb{N}$
// = number of measures on each page
// (in no. 4 the start of a new page is marked instead
// by an arbitrary event in the dedicated voice "`pg`".)
• `signumFinis` : String
// = date signature, indicates end of score

timeless parameters per voice:
• `addenda` : {**`claves`**, **`arcus`**, **`arcusMult`**}
// = type of the voice, thus the allowed event modifiers
• `cumLinea` : String
// = name of the voice with which the staff is shared
• `nomen.longum/breve` : String
// = the staff label for the first page / the subsequent pages
// (either these or `cumLinea` may be given.)
• `subordinatum`
// = marks the "piano harmonic" staves for special graphics

Table 3. Timeless parameters defined by the `tscore` "Projection_n" meta-model

it could easily be shown that Neoscore can deal with all five pieces. Nevertheless it addresses specially no. 2, and since all pieces differ also graphically (see again Table 1), a more versatile but ad-hoc rendering (393 loc) has been programmed in Java. This has been used to explore the rules and exceptions of the graphical design systematically.

Which properties must be included and which may be neglected is the central question when remodelling the graphical appearance of the scores by Feldman—as already discussed above for the end notes. For instance, Wiener regrets "[daß] die späten Werke [...] nicht mehr als Kopie der Feldmanschen Handschrift in den Handel kommen, sondern ein [...] anonymer Notensatz sie sowohl der [Seiten-] Umbrüche als auch der so wichtigen äquidistanten Taktstriche beraubt, kann nicht mehr nur ≫bedauerlich≪ genannt werden, denn hier ist eine ganze Sinnebene, vielleicht die tragende dieser Werke, besinnungslos zerstört worden." [11] [16]

The most significant issue which popped up concerns the *horizontal limiting lines*. As indicated graphically in the prefaces, the events in the high register are indicated by boxes "hanging from" the top line of their "measure box", the low register events "stands on" the bottom line, and the middle register events stand away from both. In most cases the horizontal lines in the score appear accordingly: the bottom measure box line with low register events, the top with the high register, and both with middle register events. [12]

In the original scores, Feldman adhered strictly to this "preface rule" only in no. 4. In no. 3 and 5 there is no gap between the staves, and all horizontal limiters extend over all measures (except for the middle page of no. 3 and the second half of no. 5). In nos. 1 and 3 the appearance of the horizontal limiters basically follows that preface rule, but in detail is irregular and seemingly arbitrary! In Figure 1 all horizontal lines are marked which are superfluous or missing according to that rule. [13]

The appearance of these horizontal limiting lines does not carry any semantics for the execution of the piece. Therefore our data model does not contain this information, and the implemented Java rendering algorithm strictly follows the preface rule, as the manuscript of no. 4 does. [14]

One possible interpretation could assign them *ergonomic use*, to make the synchronization of voices and/or the pitch registers more easily readable. [15] This indeed maybe the case in some particular score situations.

But seen as a whole, a different aspect seems more plausible: Knowing about the composer's interest in contemporary art, and his close contact and collaborations with Guston, Pollock, Rothko, Rauschenberg, and other painters, it is even more likely to give the graphical appearance an *aesthetic value on its own.* [16] "Painting and music [—] using them interchangeably or metaphorically in such an extent that the two art forms became one in much of Feldman's writing." [4, p. XIX] The horizontal and vertical lines can thus be seen as a direct echo of New York City's vertical and horizontal structures. [17] Seen from the top of the Empire State Building when sun sets, the selectively lit windows around follow closely the patterns of Feldman's event boxes.

6. SOUND RENDERING OF THE MODELS – AUTOMATED PERFORMANCE

6.1 General Considerations

The prefaces of all five pieces make no prescription at all about the pitches to be played. "What Feldman is assuming [...] is that the performer is a sensitive and inspired musician who has the best interests of the work at heart." [8, p. 217] These five pieces are doubtlessly among the very first in history in which pitch is not exactly specified by the composer, while onset, duration and dynamics are. John Cage, Earle Brown, and others worked at the same time in similar directions, and the true historic priorities will probably never be cleared. [18]

One possible way of interpreting the scores nowadays is an automated realization with today's computer technology. From this we expect insights in at least two con-

[11] "That the late works are no longer sold as copies of the manuscript, but in an anonymous computer rendering (which removes the page breaks and the important equidistance of the measure bars) is more than just ≫regrettable≪, because a whole layer of meaning, perhaps the fundamental one, has senselessly been destroyed." For a similar standpoint see [10, p. 4, 19].

[12] Cline states that the vertical location "is always apparent, without recourse to a ruler", means: to measuring. [3, p. 76]. Indeed it is more, namely NOTA.NONQUANTIBUS = purely syntactic = resistant to any deformation.

[13] According to [3, p. 22, 34], John Cage was involved in finding the final appearance.

[14] Two derivations from that preface rule are not taken over by our implementation:

- The larger gap and the break of the measure bars between violin and piano in no. 4.
- The aligned start of all staves and the uninterrupted horizontal staff lines (with apparently irregular exceptions) in nos. 3 and 5.

[15] Property NOTA.FACTUM.LEGERESIMPLIFICANS in [6, p. 18].

[16] Property NOTA.UTPICTURA.SIGNIFICANS in [6, p. 534].

[17] For grids as emblematic for modernism see [3, p.80,81].

[18] "John Cage was primarily interested in those aspects of Feldman's music that he understood as 'indeterminate'."[2, p. 12] Cline lists American predecessors who composed indeterminate works and discusses a possible precedence of Earl Brown [3, p. 3, 15, 39].

```
PARS sola
visuum.tactiInPagina = 10   visuum.interVoces = 0
signumFinis  = "Jan 3, 1951"
VOX fl                  nomen.longum = "Flute"   nomen.breve = "Fl."
VOX tr                  nomen.longum = "Trumpet" nomen.breve = "Trp."
VOX vl addenda = arcus  nomen.longum = "Violin"  nomen.breve = "Vl."
VOX vc addenda = arcus  nomen.longum = "Cello"   nomen.breve = "Vc."
VOX p1 addenda = claves nomen.longum = "Piano"   nomen.breve = "Pn."
VOX p2 addenda = claves cumLinea = p1
VOX ph addenda = claves nomen.longum = "◇"       nomen.breve = "◇"   subordinatum = est

T      1     2       3           4         5        6        7      8         9      10       11
VOX fl %             %.      h   -.     %           m .%     %.   l %         %.   l -   %
VOX tr %  m  -  .%   l -  %  h   %                           %.   h %         %.   l -   %
VOX vl %             % hh %  %   %.     m  - .%                     %.     mp %.   m -   %
VOX vc %             % %  mp %   %  % l %  %        % l hh % %      % % mp %  %.   m - - h  %
VOX p1 %     m1 .%   % %  l1 h7  %         % l2 % %          %.  m3 %              %.       l1
VOX p2 %
VOX ph % l5. -       -    %                         %   m1   -.  %
```

Figure 4. Timeless parameters and the first page of Feldman, *Projection 2*, as `tscore` source

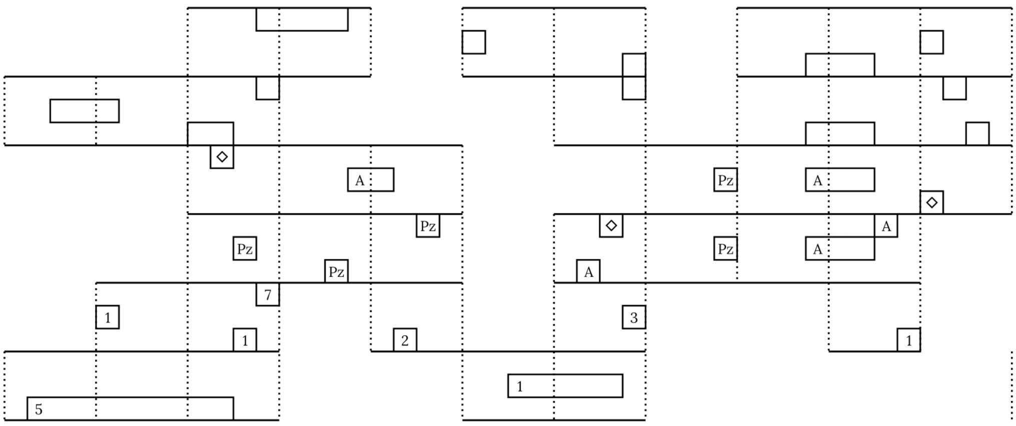


Figure 5. The first page of Feldman, *Projection 2*: data by `tscore`, rendered by Neoscore [15].

cerns: (a) Programming an algorithmic interpretation requires to take this "ancient" notation seriously—replacing the former improvisations by computer-based synthesis techniques is a kind of mathematical re-modeling. It will bring to light semantical problems and peculiarities which obviously the actors in the 1950s were not aware of, or did not care about. So our approach is a kind of "re-enactment" or "experimental archeology".

(b) Producing series of these realizations with different parameter settings can be used for empirical experiments about the effect on the listeners—a promising plan, but outside the current scope of our project.

With automated interpretation, the "sensitive performer" as requested by O'Hara is replaced by an algorithm which makes sensible transformations of the raw data delivered by an "insensible" random generator. For mimicking the improvisational behavior of "sensitive performers", the following aspects must be discussed and decided:

(A) Pitch class reservoir: A natural way for an ensemble of improvising musicians is to agree on a set of pitch classes to use throughout the performance. Since all pieces (beside no. 1) employ pianos, to play in the equally tempered tuning suggests itself. Anyway, it was the tuning system of the works of Webern and other composers of dodecaphonic serialism, the avant-garde in those times took

"onetone"—pitch class "c" only
"twotones"—pitch classes "c" and "d♭"
"bach"—pitch classes "a", "b♭, "b♮", and "c"
"tonic"— pitch classes "c", "e", and "g"
"blueNote— pitch classes "c", "e♭", "e♮", and "g"
"dominant"—pitch classes "c", "e", "g", and "b♭"
"pentatonic"—"d♭", "e♭", "g♭", "a♭", and "b♭"
"hypochromatic"— "c", "d", "e", "f♯", "g♯", and "b♭"
"dodecaphonic"—all twelve pitch classes

Table 4. Characteristic pre-selections of allowed pitch classes

as their starting point. [19] Beyond this, further restricting the set of allowed pitch classes can lead to quite different stylistic and atmospheric results—see Table 4 for typical examples. This set is defined as `PCStart` in Table 5.

(B) Definition of the three instrumental registers: In spite of their role as the fundamental datatype for the score constructions, the explanatory forewords are widely unspecific about the three instrumental registers:

(B1) It is not discussed how the registers shall be received by the listener: Shall the separation be recognized clearly

[19] For Cages's and Feldman's "mutual enthusiasm for Webern" see [5, p. 461].

as such? Or shall it only organize diffuse processes between lower and higher regions, which are perceived as a continuum? The range definitions are contained in Table 5 as the data $\texttt{range}(i)$.

The two integers in $\texttt{gaps}(i)$ are derived. Making registers explicit requires a considerable high positive integer therein. But the forewords do not even forbid *overlapping* registers = negative `gaps` values! We follow the first interpretative opinion and impose on our realizations the predicate $\texttt{wellSeparated}(i)$, with $i = 5$ as a mere matter of taste.

(B2) Naturally the high register of a lower instrument overlaps with the low registers of a higher one, w.r.t. absolute pitches, e.g. between violin and violoncello. Contrarily the lower registers in any instrument have, for physical reasons, a broader spectrum and a more sonorous timbre. By this, a low violin tone can be received as more related to a low cello tone than to a high one.

Again it is not explicitly said which alternative is intended, or whether there is any intention at all.

(B3) The ease of defining registers varies much with the kind of instrument. The predicate $\texttt{complete}(i)$ says that all pitch classes from `PCStart` are represented in every register defined for instrument i. Together with the predicate $\texttt{wellSeparated}(i)$ and the pitch class **dodecaphonic**, this requires a range of about four octaves—easily realizable by the piano, but hardly by any other instrument.

(C) Entropy and repetition:

Feldman reports about a performance in which "the players decide together, before the concert, actually to sabotage it – and they decided in this particular section they were going to play 'Yankee Doodle,' with the amount of notes called for and in the register in the score." [3, p. 45].

In general, Cline conjectures Feldman's "intense dislike for what he regarded as formulaic choices, reflecting conditioned responses or remembered sequences, [...] playing of scales or familiar [...] spoiled the performance for him."

When trying to formalize and automate these effects, both "sabotage" and "remembered sequences" can be seen as a *lack of entropy*, and a natural remedy is the ban of repetitions, similar to dodecaphonic techniques.

Those (as invented and propagated by Schoenberg, and even more in the very concentrated form practiced by Webern) were still the most recent avant-garde revolution when the *Projections* were composed. They were ideal and inspiration, but at the same time also already a tradition which Feldman and his colleagues tried to overcome.

Therefore it seems sensible to explore mathematically, how a combination of those older principles with the newly proposed register notation will work, for simulating the required "sensitive performer". For each of our realizations one of these rules must be selected:

$\texttt{eval}_{\text{Free}}$ = no restriction on repetitions of pitch classes

$\texttt{eval}_{\text{Instr}}$ = repetition not earlier than unavoidable, decided locally to each voice

$\texttt{eval}_{\text{Impro}}$ = the same, but decided globally for the whole ensemble, by improvisation

$\texttt{eval}_{\text{Comp}}$ = the same, but decided out-of-time, by "composing"

The rules $\texttt{eval}_{\text{Instr}}$ thru $\texttt{eval}_{\text{Comp}}$ correspond roughly to Schoenberg's "Wiederholungsverbot" (ban of repetition), which directly led to the invention of the "Reihentechnik" (dodecaphonic serialism). More precisely: When the interpretation starts, the selection of pitches aims to produce a *maximum-length segment without repeating pitch classes*. As soon as such a repetition becomes *unavoidable* (for what reason ever, see below!) a new such cycle is started. That means that with four elements, the initial cycle a–b–c–d can be followed by d–c–b–a, or by any other permutation.

This concretization of the rules $\texttt{eval}_{\text{Free}}$ thru $\texttt{eval}_{\text{Comp}}$ is sensible aesthetically, but also practically: With some training a musician can memorize which pitch classes from `PCStart` have been present in the current cycle (= played by themselves and maybe heard from the others) and when a new such cycle must begin.

However, the difference between $\texttt{eval}_{\text{Impro}}$ and $\texttt{eval}_{\text{Comp}}$ is significant: Whenever multiple players have to play events simultaneously, they cannot avoid possible doublings of pitch classes. [20] Therefore $\texttt{eval}_{\text{Impro}}$ and $\texttt{eval}_{\text{Comp}}$ are fundamentally different: $\texttt{eval}_{\text{Comp}}$ means "composing", when the translation from the graphical score into particular pitches happens "out of time", with full knowledge of the synchronous events. [21]

Furthermore, the key presses in the "harmonic" voices of the pianos cannot be heard by the other players, at least not immediately. Therefore we exclude them from our four pitch selection algorithms. [22]

6.2 Experimental Implementation

A main issue when realizing the pieces and defining the automated algorithm lies in the fact that during execution, aspects (A), (B), and (C) as described above are not independent but contrarily closely related: For instance, when an instrument shall play from a particular register and all pitch classes contained therein have already appeared, a new cycle must be started, as described above, in spite of still available pitch classes in other registers.

Table 5 shows the basic data types and table 6 the four different algorithms for the automated realization of the scores. `O` stands for the octave registers, in MIDI notation

[20] Of course one could establish additional communication channels between the players, beyond just playing and listening. This is not considered in this article.

[21] Such transcribing corresponds to the practical setting in Cage's *Music for Carillon (Graph) No. 1* [3, p. 37]. Further methods are sensible, e.g. that all players agree on one particular "series" = permutation of `PCStart`, which underlies the methods $\texttt{eval}_{\text{Instr}}$ to $\texttt{eval}_{\text{Comp}}$, either only verbatim or in its four modi (with retrograde, inverse, and retrograde-inverse). Or a series of such series may be defined. Or every player selects their own such series.
Or the set `PCStart` itself may change over the duration of the improvisation, which would also make the simple rule $\texttt{eval}_{\text{Free}}$ more interesting.
All these variants may be left to future work.

[22] We must exclude them from the audible realization anyhow, as long as this is restricted to third-party off-the-shelf "general MIDI" instruments. Nor can "sul ponticello" be realized. Anyhow, also in real-world execution the "sympathetic resonances [are] often inaudible". [3, p. 21; detailed discussion p. 127]. Vigil even considers "more [...] theatrical than musical reasons." [1, p. 236]

// universal types and data:
$\mathsf{O} = \{-1, 0, \ldots, 8\}$ *// = octave registers*
$\mathsf{C} = \{0, \ldots, 11\}$ *// = all (enharmonic) pitch classes*
$\mathsf{P} = \mathsf{O} \times \mathsf{C}$ *// = all pitches*
$\mathtt{key} : \mathsf{P} \to \mathbb{N}$ *// = pitch to MIDI key*
$\mathtt{key}(o, c) = c + o * 12 + 12$
$p_1 < p_2 \iff \mathtt{key}(p_1) < \mathtt{key}(p_2)$
$\mathtt{random}[X] : \mathbb{P}X \to X$ *// = select froam a non-empty set*
$\mathsf{R} = \{\mathtt{lo}, \mathtt{mid}, \mathtt{hi}\}$ *// = the three ranges low, mid, high*

// for one particular composition:
$\mathtt{I}$ *// = all involved instruments*
$\mathsf{T} = (\mathtt{I} \times \mathsf{R}) \nrightarrow \mathbb{N}_1$ *// = all data at one timepoint*
$\mathtt{score} : \mathsf{seq}\,\mathsf{T}$ *// Indexes stand for are all timepoints*
// with at least one event.

// for one particular realization:
$\mathtt{PCStart} : \mathbb{P}\mathsf{C}$ *// = selected pitch classes*
$\mathtt{range} : (\mathtt{I} \times \mathsf{R}) \to (\mathsf{P} \times \mathsf{P})$
$\mathtt{range}(i, r) = (a, b) \Longrightarrow a \leq b$
// = definitions of the three registers per instrument

// derived types and data:

$$\frac{\mathtt{range}(i, \mathtt{lo}) = (\_, a) \quad \mathtt{range}(i, \mathtt{mid}) = (b, c) \quad \mathtt{range}(i, \mathtt{hi}) = (d, \_)}{\mathtt{gaps}(i) = (\mathtt{key}(b) - \mathtt{key}(a), \mathtt{key}(d) - \mathtt{key}(c))}$$

$$\frac{\mathtt{gaps}(i) = (g, h)}{\mathtt{wellSeparated}(i) \iff g \geq 5 \land h \geq 5}$$

$\mathtt{range}' : (\mathtt{I} \times \mathsf{R}) \to \mathbb{P}\mathsf{P}$
$\mathtt{range}'(i, r) = \{\pi_1(\mathtt{range}(i, r)), \ldots, \pi_2(\mathtt{range}(i, r))\}$
$\mathtt{complete}(i) \iff \forall r \in R \bullet \mathtt{PCStart} \subseteq \pi_2 (\!| \mathtt{range}'(i, r) |\!)$

$\mathtt{choose}_{\mathrm{F/C}} : (\mathbb{P}\mathsf{P} \times \mathbb{N}) \to \mathbb{P}\mathsf{P}$

$$\frac{\mathtt{random}(A) = a \quad n > 0}{\begin{array}{l}\mathtt{choose}_{\mathrm{F}}(A, n) = \{a\} \cup \mathtt{choose}(A \setminus \{a\}, n-1) \\ \mathtt{choose}_{\mathrm{C}}(A, n) = \{a\} \\ \qquad \cup\, \mathtt{choose}(A \setminus (\mathsf{O} \times \{\pi_2(a)\}), n-1)\end{array}}$$

$\mathtt{choose}_{\_}(\_, 0) = \varnothing$
$\mathtt{choose}_{\_}(\varnothing, n) = \varnothing$
// = here an error should be reported, see text.

Table 5. Basic definitions to evaluate the "Projection_n" scores.

"4 = middle c". C are all pitch classes, `PCStart` those selected for the current interpretation. We define C by "zero stands for c" and each octave register extends from an instance of c upward. [23]

The `score` data type represents the syntactic form of the scores written by Feldman: It is a sequence, where the index stands for all those timepoints which do carry an event. Each value is of type $\mathsf{T} = (\mathtt{I} \times \mathsf{R}) \nrightarrow \mathbb{N}_1$ and maps instruments and registers to the required number of pitches to play. (With most instruments this is implicitly $= 1$, but the piano voices have larger numbers, and so has the solo violin in no. 4.)

The central issue when restricting pitch class repetition is the conflict between the still allowed pitch classes in the course of the execution, according to the selected overall strategy (represented by the set `pool`), and the ranges for the three registers per instrument (`range`). [24] The ancillary data $\mathtt{range}'$ contains the defined ranges as sets of pitches;

[23] Be aware that both facts appear natural to the contemporary reader, but are in no way necessary. [17, p. 23], according to [6, p. 327]
[24] For `range`, `gaps`, and `wellSeparated` see above; for `pool` below.

$\mathtt{eval}_{\mathrm{Free/Instr/Impr/Comp}} : \mathsf{seq}\,\mathsf{T} \to \mathsf{seq}(\mathtt{I} \leftrightarrow \mathsf{P})$

$\mathsf{T}' = (\mathtt{I} \times \mathsf{R}) \to \mathbb{N}$
$\mathtt{score}' : \mathsf{seq}\,\mathsf{T}'$
$\mathtt{score}' = ((\mathsf{dom}\,\mathtt{score} \times \mathtt{I} \times \mathsf{R}) \times \{0\}) \oplus \mathtt{score}$

$\mathtt{pool} : (\mathtt{I} \times \mathsf{R} \times \mathbb{P}\mathsf{C}) \to \mathbb{P}\mathsf{P}$
$\mathtt{pool}(i, r, C) = \mathtt{range}'(i, r) \cap (\mathsf{O} \times C)$
// = pitches currently available for playing

$\boxed{\mathtt{eval}_{\mathrm{Free}}(\mathtt{score})} = \mathtt{evF}(\!| \mathtt{score}' |\!)$
$\mathtt{evF} : \mathsf{T}' \to (\mathtt{I} \leftrightarrow \mathsf{P})$
$\mathtt{evF}(T) = \bigcup i \in I$
$\quad \bullet \{i\} \times \bigcup r \in R \bullet \mathtt{choose}_{\mathrm{F}}(\mathtt{range}'(i, r), T(i, r))$

$\boxed{\mathtt{eval}_{\mathrm{Instr}}(\mathtt{score})} = \bigcup i \in \mathtt{I}$
$\quad \bullet\, \mathtt{eval}_{\mathrm{M}}(\mathtt{PCStart}, \mathtt{extract}(i))$
$\boxed{\mathtt{eval}_{\mathrm{Comp}}(\mathtt{score})} = \mathtt{eval}_{\mathrm{M}}(\mathtt{PCStart}, \mathtt{score})$

$\mathtt{extract}(i) = \lambda(n \mapsto x) \bullet n \mapsto x \cap (\{i\} \times \mathsf{R} \times \mathbb{N})\ (\!| \mathtt{score} |\!)$
$\mathtt{eval}_{\mathrm{M/J}} : (\mathbb{P}\mathsf{C} \times \mathsf{seq}\,\mathsf{T}) \to \mathsf{seq}(\mathtt{I} \leftrightarrow \mathsf{P})$
$\mathtt{evC} : (\mathbb{P}\mathsf{C} \times \mathsf{T}) \to (\mathbb{P}\mathsf{C} \times (\mathtt{I} \leftrightarrow \mathsf{P}))$

$$\frac{\mathtt{evC}(C, x) = (C', R')}{\boxed{\mathtt{eval}_{\mathrm{M}}(C, x \blacktriangleright \alpha)} = R' \blacktriangleright \mathtt{eval}_{\mathrm{M}}(C', \alpha)}$$

$$\frac{\begin{array}{c} p = \mathtt{pool}(i, r, C) \quad c = \pi_2(\!| p |\!) \\ n_1 = \# c \quad n_2 < \# c \quad n_3 > \# c \\ p_2 = \mathtt{choose}_{\mathrm{C}}(p, n_2) \\ p_3 = \mathtt{choose}_{\mathrm{C}}(\mathtt{pool}(i, r, \mathtt{PCStart} \setminus c), n_3 - \# c) \\ \mathtt{takeAll} = \{x : c \bullet \mathtt{choose}_F(\mathtt{pool}(i, r, \{x\}), 1)\} \\ \mathtt{evC}(C', \alpha') = (C'', R'') \end{array}}{\begin{array}{l} \textbf{if } \alpha = \{(i, r) \mapsto n_1\} \cup \alpha' \\ \quad \textbf{then } C' = C \setminus c \\ \qquad \land \boxed{\mathtt{evC}(C, \alpha)} = (C'', (\{i\} \times \mathtt{takeAll}) \cup R'') \\ \textbf{else if } \alpha = \{(i, r) \mapsto n_2\} \cup \alpha' \\ \quad \textbf{then } C' = C \setminus \pi_2(\!| p_2 |\!) \\ \qquad \land \boxed{\mathtt{evC}(C, \alpha)} = (C'', (\{i\} \times p_2) \cup R'') \\ \textbf{else if } \alpha = \{(i, r) \mapsto n_3\} \cup \alpha' \\ \quad \textbf{then } C' = \mathtt{PCStart} \setminus (\pi_2(\!| p_3 |\!)) \\ \qquad \land \boxed{\mathtt{evC}(C, \alpha)} = (C'', (\{i\} \times (p_3 \cup \mathtt{takeAll})) \cup R'') \\ \textbf{else } \alpha = \varnothing \ \land \ \boxed{\mathtt{evC}(C, \alpha)} = (C, \varnothing) \end{array}}$$

$\boxed{\mathtt{eval}_{\mathrm{Impro}}(\mathtt{score})} = \mathtt{eval}_{\mathrm{J}}(\mathtt{PCStart}, \mathtt{score})$

$$\frac{\begin{array}{c} (\_, P_i) = \mathtt{evC}(C, x \cap (\{i\} \times R \times \mathbb{N})) \\ P = \bigcup_{i \in \mathtt{I}} P_i \\ C_{\mathrm{P}} = \pi_2(\!| \pi_2(\!| P |\!) |\!) \\ C' = \textbf{if } C_{\mathrm{P}} \subset C \textbf{ then } C \setminus C_{\mathrm{P}} \textbf{ else } \mathtt{PCStart} \setminus (C_{\mathrm{P}} \setminus C) \end{array}}{\boxed{\mathtt{eval}_{\mathrm{J}}(C, x \blacktriangleright \alpha)} = P \blacktriangleright \mathtt{eval}_{\mathrm{J}}(C', \alpha)}$$

Table 6. The implemented four different ways of evaluating the "Projection_n" scores. Terms defined are framed.

`complete()` is the fact that every such range contains all selected pitch classes—this is not mandatory!

The $\mathtt{choose}_{\_}(A, n)$ function selects from the given set a given number of elements: $\mathtt{choose}_F$ removes only the selected pitch, but $\mathtt{choose}_C$ all instances of the selected pitch class. [25]

[25] As implemented in Table 6. the code may call the function `random(..)` redundantly, e.g. to select one element from a one element

Table 6: The result of each variant of $\texttt{eval}_{\_}(..)$ is always a sequence of values (at the same timepoints as the input data) of type $\mathsf{T} \times \mathsf{P}$, which map instruments to sets of pitches.

(The transformed data T' and $\texttt{score}'$ are for technical reasons only and replace any undefined coordinate in the score data by an explicit zero.)

$\texttt{pool}(..)$ calculates permitted pitches by combining the range restriction for an instrument and its register with a set of pitch classes.

The ancillary functions $\texttt{evF}(..)$ and $\texttt{evC}(..)$ process all score data at one particular timepoint.

$\texttt{eval}_{\mathrm{Free}}$ simply steps through the timepoints and applies to them $\texttt{evF}(..)$, the random choice for the required number of pitches, without restrictions.

Whenever the set returned by $\texttt{choose}_{\_}(..)$ is smaller than the number requested, this is because the particular range has been erroneously defined when preparing the interpretation. It can be checked in advance for all combinations of score position, instrument, and register that the requested number of pitches to play does not exceed the number of playable pitches:

$$\forall n : \mathbb{N}, i : \texttt{I}, r : \mathsf{R} \\ \bullet\ \texttt{score}(n)(i, r) \leq \#\texttt{pool}(i, r, \texttt{PCStart})$$

This also applies for the other evaluation modes.

$\texttt{eval}_{\mathrm{Instr}}$ and $\texttt{eval}_{\mathrm{Comp}}$ both use the ancillary function $\texttt{eval}_{\mathrm{M}}$, which simply maps $\texttt{evC}$ over the timepoints of the score: The resulting pc set from one timepoint is the input for the processing of its successor.

With $\texttt{eval}_{\mathrm{Comp}}$ this is done with the complete score at once; with $\texttt{eval}_{\mathrm{Instr}}$ for each instrument with a filtered score $\texttt{extract}(i)$ separately, merging the results. The function $\texttt{evC}(..)$ realizes the repetition restriction at one particular timepoint: Its input are the score data of type $\mathsf{T} = (\texttt{I} \times \mathsf{R}) \nrightarrow \mathbb{N}$ and the set of currently allowed pitch classes $\mathbb{P}\mathsf{C}$. It delivers an assignment of pitches $\texttt{I} \times \mathsf{P}$ and the set of allowed pitch classes for the subsequent timepoint.

Its implementation is *heuristic*: The resource assignment problem as such is NP-hard, but a heuristic approach seems more appropriate and better comprehensible:

First, all combinations of events and registers for which the number of still available pitch classes is *identical* to the number of required pitches ($n_1 = \# c$; the projection π_2 maps a pitch to its pitch class) get the events they need. This is because these assignments must be done in *any* solution which does not start a new pitch class cycle. After the assignment, the function is called again with input data cut down to α' and the set of available pitch classes readjusted to C'.

Please note that the ancillary function takeAll does call the choose operation separately for each pitch class. Calling choose directly for all pitch classes would deliver the same result, but maybe make superfluous calls to random: For $\texttt{pool}(..) = \{\texttt{C3}, \texttt{C4}, \texttt{D3}\}$ only one decision is required, namely between the two representatives of pitch class c, but the first of direct random choices could deliver the D3.

If no such event is found, a pair with some degree of freedom and without the need of starting a new cycle is processed ($n_2 < \# c$). This is the point where the algorithm gets heuristic, because collecting all such events and distributing the available pitch classes among them with more global knowledge could bring better results. But in our implementation the assignment p_2 is selected locally. All events of this kind are processed in second line to get as many assignments as possible into the current repetition cycle.

In both cases the input for the recursive call to this function is cut down to the remaining events at this timepoint α' and the remaining pitch classes C'.

With least priority the pairs are treated which require a new cycle anyhow ($n_3 > \# c$): First takeAll uses up the rest of the old cycle. The remaining requests p_3 are chosen from PCStart, amd their classes are subtracted for the starting set C' for the next assignment.

$\texttt{eval}_{\mathrm{Impro}}$ is different, because it must reflect the principle that the improvising players do not know the decisions of the others before they have heard them. The function steps through the timepoints by the ancillary function $\texttt{eval}_{\texttt{J}}$. For each timepoint evC is applied to all voices *separately*, with the same input pc set C. When afterwards all assigned pitches fit into this set (P_i are the assignments calculated for the singleton set $\{i\}$; the inner π_2 maps the pairs from $\texttt{I} \leftrightarrow \mathsf{P}$ to the pitches, the outer maps these to their pitch classes) then the rest set is taken as the input for processing the subsequent timepoint. [26] Otherwise at least one of the "improvising" instruments has encountered the necessity to start a new cycle. Consequently a new cycle is started also for the whole ensemble: The complete selected set PCStart is the input for the next timepoint, after removing all pitch classes which have been assigned here but are not provided by C.

The software can be downloaded from http://bandm.eu/feldmanProjections.html. It is in the public domain; source code publication is in preparation.

7. AUTOMATED ANALYSES OF THE MODELS

As mentioned in the previous sections, we applied programmed analysis to the data model for sorting out critical properties like overlapping piano chords or string multistops.

The idea of computer based analysis of a work of Feldman has been realized by Thomas Hummel [10]. With an Atari home computer of the Nineties in the programming language Forth, statistical data mostly about pitches and pitch classes are extracted from "Untitled Composition" (1981). However, the fundamental theses and final conclusions are taken "manually".

More recently, David Cline applied statistical analyses to the "graph music of Morton Feldman", i.e. to Projection 1–5 and beyond. His results w.r.t. the "even-handed distribution" to the three registers and to the instruments

set. Any implementation should react appropriately, e.g. not protocol or analyze such a "non-random" request.

[26] Thus accidental doublings in the resulting sum are not the *unavoidable* repetitions from the cycle rule stated above.

```
                  events per register: numbers (percentage):          summed duration per register (idem):
Proj 1         harm  3;10; 7 (= 15.0; 50.0; 35.0 %)   3; 11; 11 (= 12.0; 44.0; 44.0 %)
Proj 1         pizz 18; 7;15 (= 45.0; 17.5; 37.5 %)  18;  7; 15 (= 45.0; 17.5; 37.5 %)
Proj 1         arco  8; 4; 3 (= 53.3; 26.7; 20.0 %)   8;  7;  6 (= 38.1; 33.3; 28.6 %)
Proj 1          sum 29;21;25 (= 38.7; 28.0; 33.3 %)  29; 25; 32 (= 33.7; 29.1; 37.2 %)
Proj 1            harm  20 (26.7 %)   25 (29.1 %) // = events (percentage) per voice, summed-up durations (idem)
                  pizz  40 (53.3 %)   40 (46.5 %)
                  arco  15 (20.0 %)   21 (24.4 %)
Proj 2            fl 16; 6;10 (= 50.0; 18.8; 31.3 %)  48; 10; 21 (= 60.8; 12.7; 26.6 %)
Proj 2            tr 10; 6;17 (= 30.3; 18.2; 51.5 %)  20; 17; 36 (= 27.4; 23.3; 49.3 %)
Proj 2            vl  7;10;10 (= 25.9; 37.0; 37.0 %)  11; 16; 24 (= 21.6; 31.4; 47.1 %)
Proj 2            vc 13;15;23 (= 25.5; 29.4; 45.1 %)  22; 26; 54 (= 21.6; 25.5; 52.9 %)
Proj 2         p1+p2 13; 7;12 (= 40.6; 21.9; 37.5 %)  38; 10; 31 (= 48.1; 12.7; 39.2 %)
Proj 2           sum 59;44;72 (= 33.7; 25.1; 41.1 %) 139; 79;166 (= 36.2; 20.6; 43.2 %)
Proj 2              fl  32 (18.3 %)   79 (20.6 %)
                    tr  33 (18.9 %)   73 (19.0 %)   approximately equal distribution between and intern of
                    vl  27 (15.4 %)   51 (13.3 %)   two subgroups (fl, trp, vl / vc, p) [3, p. 145]
                    vc  51 (29.1 %)  102 (26.6 %)
                 p1+p2  32 (18.3 %)   79 (20.6 %)
Proj 3     p1+p1b  1; 6;14 (=  4.8; 28.6; 66.7 %)   1; 14; 41 (=  1.8; 25.0; 73.2 %)
Proj 3     p2+p2b  7; 8; 3 (= 38.9; 44.4; 16.7 %)  17; 19;  3 (= 43.6; 48.7;  7.7 %)
Proj 3        sum  8;14;17 (= 20.5; 35.9; 43.6 %)  18; 33; 44 (= 18.9; 34.7; 46.3 %)
Proj 3         p1+p1b  21 (53.8 %)   56 (58.9 %)
               p2+p2b  18 (46.2 %)   39 (41.1 %)
Proj 4          vlH  6;12;14 (= 18.8; 37.5; 43.8 %)   8; 17; 23 (= 16.7; 35.4; 47.9 %)
Proj 4          vlP 16;10;18 (= 36.4; 22.7; 40.9 %)  16; 10; 18 (= 36.4; 22.7; 40.9 %)
Proj 4          vlA  8;14; 3 (= 32.0; 56.0; 12.0 %)  14; 28;  5 (= 29.8; 59.6; 10.6 %)
Proj 4        p+p2 30;26;30 (= 34.9; 30.2; 34.9 %)  34; 41; 36 (= 30.6; 36.9; 32.4 %)
Proj 4         sum 60;62;65 (= 32.1; 33.2; 34.8 %)  72; 96; 82 (= 28.8; 38.4; 32.8 %)
Proj 4            vlH  32 (17.1 %)   48 (19.2 %)
                  vlP  44 (23.5 %)   44 (17.6 %)
                  vlA  25 (13.4 %)   47 (18.8 %)   "even-handed distribution" [3, p. 140]
                 p+p2  86 (46.0 %)  111 (44.4 %)

...
Projection 4, events by duration:
dura =  1  count =131  (74.4 %)  [1, 2, 5, ...
dura =  2  count = 15  ( 8.5 %)  [224, 193, 290, 68, 101, 293, 137, 202, 205, 17, 274, 149, ...
dura =  3  count = 11  ( 6.3 %)  [256, 288, 36, 132, 170, 267, 299, 141, 13, 207, 239]
dura =  4  count =  4  ( 2.3 %)  [67, 281, 73, 74]
dura =  5  count =  1  ( 0.6 %)  [324]   "a duration he had not yet used" [1, p. 255]
...
Projection 1 --- recognized regions:
sync instruments :
sounding instruments, exact :
thickness 0  length 1: [9, 12, 31, 37, 50, 53, 55, 67, 77, 86, 92, 106, 108, 110, 152, 178]
thickness 0  length 2: [14, 25, 28, 34, 39, 47, 57, 63, 69, 89, 98, 112, 118, 132, 155, 168, ...
thickness 0  length 3: [0, 4, 17, 43, 72, 94, 102, 121, 135, 148, 158, 181]
thickness 0  length 4: [81, 125, 140, 162]
thickness 0  length 11: [188]       "the work's most striking aspect is the longest silence" [3, p. 18]
...
Projection 2 --- recognized regions:
sync instruments :
    2:  [10, 40, 69, 83, 85, 218, 224, 227, 277, 280, 309, 322, 326]
    3:  [11, 27, 58, 61, 79, 228]
    4:  [35, 91, 96, 103, 204]
    5:  [171]
sync keys cross instruments :
    2:  [10, 40, 69, 83, 85, 218, 224, 227, 277, 280, 309, 322]
    3:  [58, 61, 79, 228]
    4:  [35, 91, 96, 103, 204]
    5:  [27, 171]
    9:  [11]
   13:  [326]
sounding instruments, exact :
thickness 0  length 1: [19, 59, 78, 84, 86, 90, 95, 145, 147, 151, 158, 160, 162, 188, 194, 255, ...
thickness 0  length 2: [0, 28, 43, 92, 140, 166, 169, 199, 202, 230, 275, 278, 283]
thickness 0  length 3: [5, 32, 80, 119, 153, 190, 296]
thickness 0  length 4: [21, 46, 53, 65, 99, 213, 244]
thickness 0  length 5: [72, 205, 290]
thickness 0  length 6: [303]
thickness 0  length 9: [104]        "the longest [pause] occurs just after the midpoint." [3, p. 19]
thickness 0  length 12: [175]
thickness 1  length 1: [8, 20, 45, 57, 60, 64, 77, 94, 113, 146, 152, 159, 161, 168, 187, 189, 193, ...
thickness 1  length 2: [2, 12, 25, 30, 38, 41, 70, 156, 225, 263]
thickness 1  length 3: [50, 87, 142, 148, 163, 258, 266, 270, 287]
thickness 1  length 4: [15, 195, 299]
thickness 1  length 7: [237, 248]
thickness 1  length 18: [122]
thickness 2  length 1: [4, 14, 40, 69, 83, 85, 114, 211, 222, 227, 236, 273, 277, 321, 324]
...
thickness 5  length 1: [171, 174, 220, 336]
thickness 5  length 3: [332]        "In the last few seconds, all the instruments play simultaneously." [3, p. 20]
```

Table 7. Parts of the statistical data extracted from the `tscore` model, and its relations to statements by Cline [3] and Vigil [1]. (All start-points given as zero-based "ictus" = quarters.)

```
Projection 2 -- sounding instruments, minimally :
thickness 1  length 1: [20, 45, 77, 79, 83, 85, 91, 94, 103, 146, 152, 159, 161, 168, 187, ...
thickness 1  length 2: [30, 57, 156]
thickness 1  length 3: [2, 25, 50, 69, 87, 96, 142, 148, 163, 210, 266, 280, 287]
thickness 1  length 4: [171, 195, 299]
thickness 1  length 5: [60, 270]
thickness 1  length 6: [113]
thickness 1  length 7: [248, 258]
thickness 1  length 8: [35]
thickness 1  length 11: [8]
thickness 1  length 12: [232]
thickness 1  length 13: [217]
thickness 1  length 18: [122]
thickness 1  length 28: [309]     "The most impressive [staggered overlapping sequence is] at the very end of the work." [3, p. 19]
...
Projection 4 --- recognized regions:
sync instruments :
    2:  [9, 22, 83, 87, 98, 108, 111, 129, 132, 141, 149, 164, 170, 193, 207, 211, 222, 251, ...
sync keys cross instruments :
    2:  [83, 108, 132, 164, 261]
    3:  [22, 111, 170, 207, 222]
    4:  [9, 141, 267]
    5:  [87, 149, 211, 251]
    8:  [129, 193]
    9:  [98]
   11:  [324]     "an extremely dense eleven note sonority." [1, p. 255]
```

Table 8. Statistical data, continued.

in Projection 4 [3, p. 140] are easily reproduced with the `tscore` data models, see Table 7. But is also shows that both properties do not appear in the other pieces.

Once these analyses have been implemented, it took few more lines of code to apply them to the sums of durations. Also here no further equilibria appear.

Speaking of "the longest silence" or "the most impressive [. . .] of staggered sequences" (p. 10, 19, etc.) transits from mere statistical to structural claims; the underlying facts are verified by our analysis.

Cline posits repeated and mirrored "modules" in the "register contours" of the score (p. 164) and "trajectories" (mere graphical phenomena because they transit the limits of the staves) in survived sketches (p. 169). Both are a possible subject of future automated analysis by pattern recognition.

Vigil proposes for an analysis of Projection 4 the notions "restriction", "exclusion", "diversity", "saturation", "density", and "novelty" [1]. It is not quite clear in how far these stand for structural/syntactic properties of the score or for psychological categories in reception. The former could be extracted automatically from the `tscore` model, after appropriate formalization, which appears as promising future work.

8. REANIMATING THE NOTATION

For further research, especially on the psychology of reading and playing by human actors, the first author composed small studies using Feldman's notation ("Drei kleine Studien nach Morton Feldman", Lepper op. el. 20).

Automated performance already showed that with additional restrictions on permitted pitch classes and repetition avoidance, even opposed styles of intention and effect can be supplied. They range from abstract serialism (with precise quasi-mathematical messages to the listener) to easy-listening soundscapes.

It turned out for the first and last of the three movements, that it was appropriate w.r.t. their content to *generate the Feldman scores by an algorithm*. This was remarkably easy (300+200 loc) and proved again the versatility of the basic design of `tscore`.

For the first movement even *score stacking* was applied: A second `tscore` format describes only the curves of the limits of the random values which control the generation of the Feldman events. This approach showed much easier handling of input data, with aesthetically satisfactory outcome.

9. COMPREHENSION

The compositions *Projection 1* to *Projection 5* by Morton Feldman from 1950/51 are important historic avant-garde compositions. They are among the first pieces which founded modern graphic notation and gave the decision about pitches to the players.

`Tscore` is a meta-meta-model model which allows to construct (in few lines of code) a meta-model which allows to create (in little editing time) computer models of all five scores very conveniently. All code is in the public domain—open source publishing is in work.

These score models have been used to reconstruct the concrete graphical appearance, to generate possible sound realizations, and to answer particular analytical questions. During the reconstruction processes of sound and graphics, difficulties became obvious which have not yet been discussed since the days of their composition.

Acknowledgments

Many thanks to Andrew Yoon, NYC, principal developer of Neoscore [15], and to the DigiScore ERC project which supported it. Yoon's work on *Projection 2* inspired our research. Thanks to the anonymous reviewers who helped to improve this article considerably.

10. REFERENCES

[1] R. Vigil, "Compositional Parameters: 'Projection 4' and an Analytical Methodology for Morton Feldman's Graphic Works," *Perspectives of New Music*, vol. 47, no. 1, pp. 233–267, 2009.

[2] A. Noble, *Composing Ambiguity: The Early Music of Morton Feldman*. Farnham, UK: Ashgate, 2013.

[3] D. Cline, *The Graph Music of Morton Feldman*. Cambridge, UK: Cambridge University Press, 2016.

[4] M. Feldman, *Give My Regards to Eighth Street – Collected Writings*. Cambridge, MA: Exact Change, 2000.

[5] B. Boutwell, "Morton Feldman's graphic notation: projections and trajectories," *Journal of the Society for American Music*, vol. 6, no. 4, pp. 457–82, 2012.

[6] M. Lepper, "De linguis musicam notare," Ph.D. dissertation, Universität Osnabrück, Osnabrück, 2021, https://www.epos.uni-osnabrueck.de/buch.html?id=150[20230920].

[7] M. Lepper and B. Trancón y Widemann, "Beam it up! – a classification grid for historic and contemporary practices of beaming by mathematical re-modelling," in *Proceedings of TENOR 2023 conference*, 2023, https://zenodo.org/record/8079306.

[8] F. O'Hara, "New directions in music: Morton Feldman," 1959, (in [4]).

[9] M. Erdmann, "Zusammenhang und Losigkeit," *Musik-Konzepte*, vol. 48/49, 1986.

[10] T. Hummel, *Morton Feldmans "Untitled Composition" (1981) für Cello und Klavier. Eine computergestützte Analyse*. Saarbrücken: Pfau, 1994.

[11] M. Lepper and B. Trancón y Widemann, "tscore: Makes computers and humans talk about time," in *Proc. KEOD 2013, 5th Intl. Conf. on Knowledge Engineering and Ontology Development*, J. Filipe and J. Dietz, Eds., instincc. Portugal: scitePress, 2013, pp. 176–183, http://markuslepper.eu/papers/tscore2013.pdf[20231003].

[12] ——, *Example Instances of the TScore Project Infrastructure*, 2013, http://markuslepper.eu/sempart/tscoreInstances.html[20230920].

[13] ——, "Translets – parsing diagnosis in small dsls, with permutation combinator and epsilon productions," in *Tagungsband des 35ten Jahrestreffens der GI-Fachgruppe "Programmiersprachen und Rechenkonzepte*, ser. IFI Reports, vol. 482. University of Oslo, 2018, pp. 114–129, http://urn.nb.no/URN:NBN:no-65294[20230920].

[14] A. Danial, "cloc – count lines of code (v1.92)," Tech. Rep., 2021, https://github.com/AlDanial/cloc [20231003]. [Online]. Available: https://doi.org/10.5281/zenodo.5760077

[15] A. Yoon, "Neoscore – notation without bars," Tech. Rep., https://github.com/DigiScore/neoscore/tree/main/examples/feldman_projection_2 [20230920].

[16] O. Wiener, *Morton Feldmans "The Viola in my Life" (1970–71)*. Saarbrücken: Pfau, 1996.

[17] H. Riemann, *Notenschrift und Notendruck. Festschrift zur 50-jährigen Jubelfeier des Bestehens der Firma Röder*. Leipzig: C.G. Röder GmbH, 1896.

[18] J. Spivey, *The Z Notation: a reference manual*. Prentice Hall, 1988, https://github.com/Spivoxity/zrm/blob/master/zrm-pub.pdf[20231003].

A. MATHEMATICAL NOTATION

The employed mathematical notation is fairly standard, inspired by the Z notation [18]. The following table lists some details:

$\mathbb{N}$	All natural numbers, incl. Zero.
$\mathbb{N}_1$	All natural numbers without Zero.
$\#A$	The cardinality of a finite set = the natural number of the elements contained.
$\mathbb{P}A$	Power set, the type of all subsets of the set A, incl. infinites.
$A \setminus B$	The set containing all elements of A which are not in B.
$A \times B$	The product type of two sets A and B, i.e. all pairs $\{c = (a, b) \vert a \in A \wedge b \in B\}$.
π_n	The nth component of a tuple.
$A \rightarrow B$	The type of the *total* functions from A to B.
$A \nrightarrow B$	The type of the *partial* functions from A to B.
$A \leftrightarrow B$	The type of the relations between A and B.
$a \mapsto b$	An element of a relation; simply another way to write (a, b)
$f \llparenthesis s \rrparenthesis$	The image of the set s under the function or relation f
$\mathsf{dom}\, A, \mathsf{ran}\, A$	Domain and range of a relation.
$r \oplus s$	Overriding of function or relation r by s. Pairs from r are shadowed by pairs from s: $r \oplus s = \big(r \setminus (\mathsf{dom}\, s \times \mathsf{ran}\, r)\big) \cup s$, with dom and ran being domain and range, resp.
$\mathsf{seq}\, A$	The type of finite sequences from elements of A, i.e. of maps $\mathbb{N} \nrightarrow A$ with a contiguous range $\{1..n\}$ as its domain. Instances are notated by listing the range elements in $\langle \ldots \rangle$.
$a \blacktriangleright \beta$	A sequence seen as a first element a and the rest sequence β (our extension). Same as $\langle a \rangle$ concatenated with β.

The frequently used notation

$$\frac{\begin{array}{c} a \\ b \qquad c \end{array}}{d}$$

means as usual $a \wedge b \wedge c \implies d$. Nearly always it should be read as an *algorithm*: "For to calculate d, try to calculate a, b and c. If this succeeds, the answer d is valid."

Functions are considered as special relations; relations as sets of pairs. So with functions, expressions like "$f \cup g$" are defined.

DJSTER REVISITED – A PROBABILISTIC MUSIC GENERATOR IN THE AGE OF MACHINE LEARNING

Georg Hajdu
ligeti center
Hamburg University of Music and Drama
georg.hajdu@hfmt-hamburg.de

ABSTRACT

DJster is a probabilistic generator for musical textures based on Clarence Barlow's legacy program Autobusk, further developed by Hajdu since 2008. The 2023 revision for Max and Ableton Live includes new features that improve the versatility of the application and enable data exchange between the synchronous and asynchronous incarnations. The synchronous incarnation of DJster can be used to preview a texture to be further developed as a sketch in the asynchronous one. DJster allows the real-time addition and modification of tonal and metric profiles departing from Barlow's original fixed-input paradigm. This motivated an exploration of metric interpolations by means of self-organizing maps and an extension of Jean-Claude Risset's illusion of an ever-accelerating rhythm. Furthermore, the implementation of a novel *melodic cohesion* parameter allows transitions from a sequence of events to a probabilistic process, the latter being the original modus operandi. Finally, DJster as a style-agnostic music generator can be embedded in machine-learning contexts to make user interaction a more rich and intuitive experience.

1. INTRODUCTION

DJster is a generative music system based on work by the composer and computer musician Clarence Barlow done between 1986 and 2000 [1] whose development was taken over by Hajdu in 2008; its genesis being outlined in a 2016 TENOR conference paper [2]. It is a sophisticated, style-agnostic system for the generation of tonal/atonal and metric/ametric music based on a stochastic principles. In comparison to "opaque" machine learning systems, DJster represents a "transparent" approach controlled by 12 distinct, yet interdependent parameters such as *event density*, *harmoniclarity* or *tonic pitch*. By adopting a mathematical method, DJster also allows the exploration of a large number of microtonal scales without requiring any pretraining of the system. DJster exists in several incarnations for Max (Figure 1) and Ableton Live. The features added recently are concerned with the ability to interpolate between complex meters and pitch sets in real time as well as passing parameter settings to a MaxScore plug-in for the generation of music notation.

In the original version of Autobusk, Barlow made metric and tonal-semblance profiles strictly dependent on the meter and scale files to be used [7]. Their hierarchies were calculated according to the algorithms he developed for this purpose. In contrast, DJster permits arbitrary, user-defined profiles which need not be strictly hierarchical. In combination with the *melodic cohesion* parameter, DJster can thus create transitions between motives and the probabilities assigned to the members of these motives.

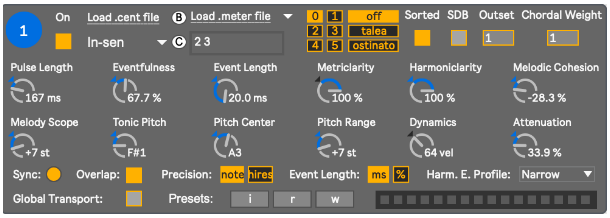


Figure 1. GUI of the 2024 version of DJster.

DJster has been used in countless projects ranging from dance-theater pieces [2] and real-time notation of EEG signals to sound interventions in hospital waiting areas [4]. Exposing all its parameters to the host environment, it can be used as event generator in Ableton Live as well as interactively in a man-machine dialog.

As outlined in [2], parameter settings could also be considered a type of notation for *musical texture*, particularly if they could be represented symbolically via shorthand notation. It requires some intimacy with the system to predict the resulting textures, but experience has shown that the connection between settings and outcome can be learned quickly and applied reliably during the composition and interaction process. Using machine learning (ML), this connection can be inverted for the generation of parameter settings via the creation of a large corpus of music generated by DJster. Upon training the ML system, music fed into it would generate settings resulting in music closely related to the input. This could spawn a recursive and potentially interesting dialog between musical input and output.

2. METER

2023 marks the year of György Ligeti's 100th birthday as well as Clarence Barlow's passing[1]. Both composers have shown great interest in meter and polymeter in their compositions and music-theoretical work. Ligeti prided himself of generalizing the medieval concept of the hemiola apparent in many a piano etude [6] while Barlow undertook quantitative research on metric hierarchy in context of his piano piece *Çoğluotobüsişletmesi* [7]. In his 1980 essay *Bus Journey to Parametron,* he pointed out the lack of available literature on this topic at the time [8]. Barlow's theory of meter is characterized by the mathematical concept of *indispensability* from which a *metric profile* can be derived for any given meter (e.g. [2, 0, 1] for triple meter). Considering prime *stratification divisors,* his formula yields profiles for complex multiplicative and additive meters, e.g. [8, 0, 3, 6, 1, 4, 7, 2, 5] for compound triple meter. While in the original version of Autobusk, the metric profiles had to be calculated and saved to an *.IDP* file, DJster generates metric profiles on the fly and accepts arbitrary *named* profiles that may be modified in real-time. This can be achieved by sending it a message consisting of the message name *append-meter*, a *meter_name* and a *list of values* for the definition of a metric profile referred to as *meter_name*. This feature facilitated the experiments on metric interpolation by means of ML Hajdu undertook during the summer of 2023, aiming to provide a framework for flexible and fluid rhythm generation.

2.1 Self-Organizing Networks

We explored the ability of *self-organizing maps* (SOM) to distinguish between the hemiolas evident in 3/4, 6/8 and 12/16 meters[2], where the metric *stratifications* consist of the permutation of the elements 2 and 3 (and powers thereof). With a Max patch created in 1992 in David Wessel's Introduction to Computer Music class at UC Berkeley, one of the participating students showed that a *multilayer perceptron* (MLP) could classify a random stream of pulses according to eight profiles which the MLP was trained to (**Table 1**). Furthermore, an *inverse* network generated an accompaniment that best fit the random series of pulses.

3x2x2:	1 0 0 0 0 0 0 0	1 0 0.56 0.1 0.82 0.27 0.64 0.18 0.91 0.36 0.73 0.45
3x2:	0 1 0 0 0 0 0 0	1 0 0.170 0 0.670 0 0.330 0 0.830 0 0.500 0
3:	0 0 1 0 0 0 0 0	1 0 0 0 0.330 0 0 0 0.670 0 0 0
2x3x2:	0 0 0 1 0 0 0 0	1 0 0.56 0.1 0.73 0.27 0.91 0.36 0.64 0.18 0.82 0.45
2x3:	0 0 0 0 1 0 0 0	1 0 0.170 0 0.500 0 0.830 0 0.330 0 0.670 0
2:	0 0 0 0 0 1 0 0	1 0 0 0 0 0 0.500 0 0 0 0 0
2x2x3:	0 0 0 0 0 0 1 0	1 0 0.36 0.73 0.18 0.56 0.91 0.1 0.45 0.82 0.27 0.64
2x2:	0 0 0 0 0 0 0 1	1 0 0 0.250 0 0 0.750 0 0 0.500 0 0

Table 1. Training data for the eight profiles used for the generation of meters with 12 pulses (inverse model). The red numbers encode indexes for the eight meters while the black numbers represent associated metric profiles, normalized to the 0–1 range.

[1] Sadly, Clarence Barlow died on June 29, 2023.

[2] We used the Max ml.som object from the ml.star package [9].

Exposing a SOM to the same profiles, the network accurately represents them in a 2-dimensional space (Figure 2) with the 2 axes indicating stratification depth (x-axis) and stratification type (y-axis). The red area denotes the division of the measure by 3 beats getting increasingly brighter as it is gradually subdivided into 12 pulses. The black area transitioning to brown (6/8) and blue (12/16) indicates the division of the measure by 2 (and then further divided into 2 and 2x3 or 3 and 3x2, totaling 12 pulses). The square is roughly divided into eight zones showing the learned profiles as well as the interpolations between them.

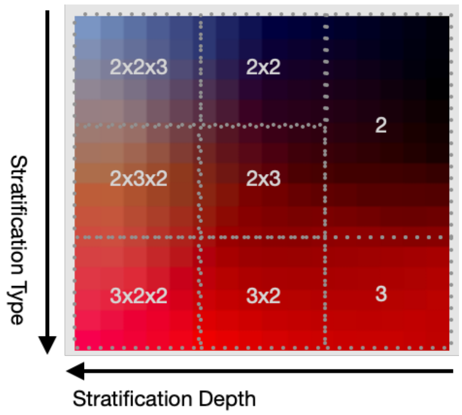


Figure 2. An ml.som object was trained to metric profiles for duple and triple meters and their subdivisions.

2.2 'Bohlen-Pierce' Meter

We then investigated how a SOM would fare if exposed to far more complex meters made of primes 3, 5 and 7. Trained to 16 different vectors with 105 values each – representing a meter made of these primes and their combinations – the SOM produced a result which was more difficult to interpret (see Figure 3). What in the previous example with 12 pulses looked like a logical arrangement in two dimensions, now seemed more like a scattering of profiles along the two axes, with some of the profiles missing. We concluded that the scattering would be the result of a dimensionality reduction performed by the SOM, for the space formed by the stratifications of these three primes may not be accurately mapped in two dimensions. To test this assumption, we calculated the Euclidian distance between the profiles and performed multidimensional scaling which yielded – in three dimensions – a sensible arrangement supporting our hypothesis. We were also able to discern some similarities to the results of the SOM (Figure 4).[3]

[3] For instance, 3 and 7 are at a closer distance to each other than 5, and the six three-strata meters are clustered in three areas (which may explain why the distinction between the pairs of meters within the clusters is blurred or why one member of a pair may have eclipsed the other).

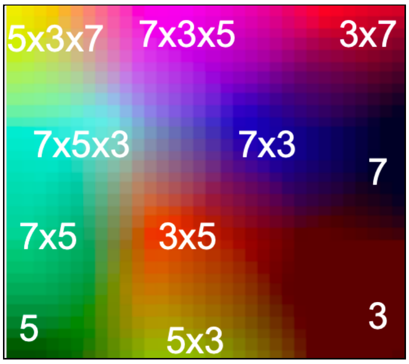


Figure 3. A SOM trained to 16 metric profiles based on primes 3, 5 and 7. In comparison to Figure 2 the zones appear scattered. This is most likely due to dimensionality reduction.

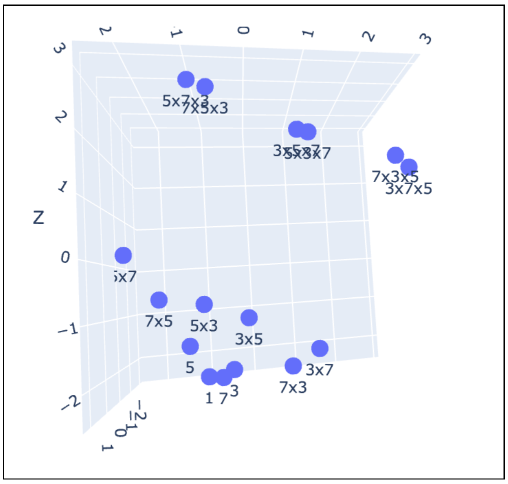


Figure 4. Multidimensional scaling of the Euclidian distances between the metric profiles in 3D, substantiating the assumption of a dimensionality reduction in the SOM.

2.3 Multilayer Perceptrons

The question arose as to how to best control the space formed by these primes with available gestural or fader controllers to provide interpolated profiles to DJster in real-time. One possibility was to use the SOM despite its seemingly "imperfect" arrangement of the profiles; it still yielded intriguing rhythmic patterns. The other was to train a multilayer perceptron to an arbitrary spatial arrangement of the profiles. We conceived of each of the six three-strata profiles as "fader" with a gradual transition from a one-pulse meter (0^{th} stratum) to meters with one, two and finally three strata, e.g. 1→3→3x5→3x5x7, corresponding to fader values 0., 0.33, 0.66 and 1 (Figure 5).

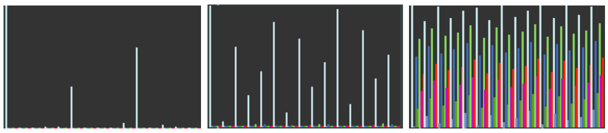


Figure 5. Metric profiles for fader values [.33, 0, 0, 0, 0, 0], [.66, 0, 0, 0, 0, 0] and [1., 0, 0, 0, 0, 0].

For this, we configured the Max-Java nnLists object to six input neurons receiving their input from the faders and 105 output neurons encoding the metric profiles. Interpolations were performed as crossfades between pairs of meters, such as 3x5x7→7x5x3 (Figure 8).

Figure 6. Notation of the interpolation between a 7x5x3 meter and a 3x5x7 meter according to interpolated metric profiles, performed by an MLP trained to 16 different metric profiles. The notehead sizes indicate the velocities of the notes which in turn are dependent on their indispensability via the *attenuation* parameter. While on the top system every third note is stressed, it is every seventh on the bottom system (also revealing the 2+2+3 nature of this additive meter). The systems in between are characterized by metric ambiguity resulting from the interpolation.

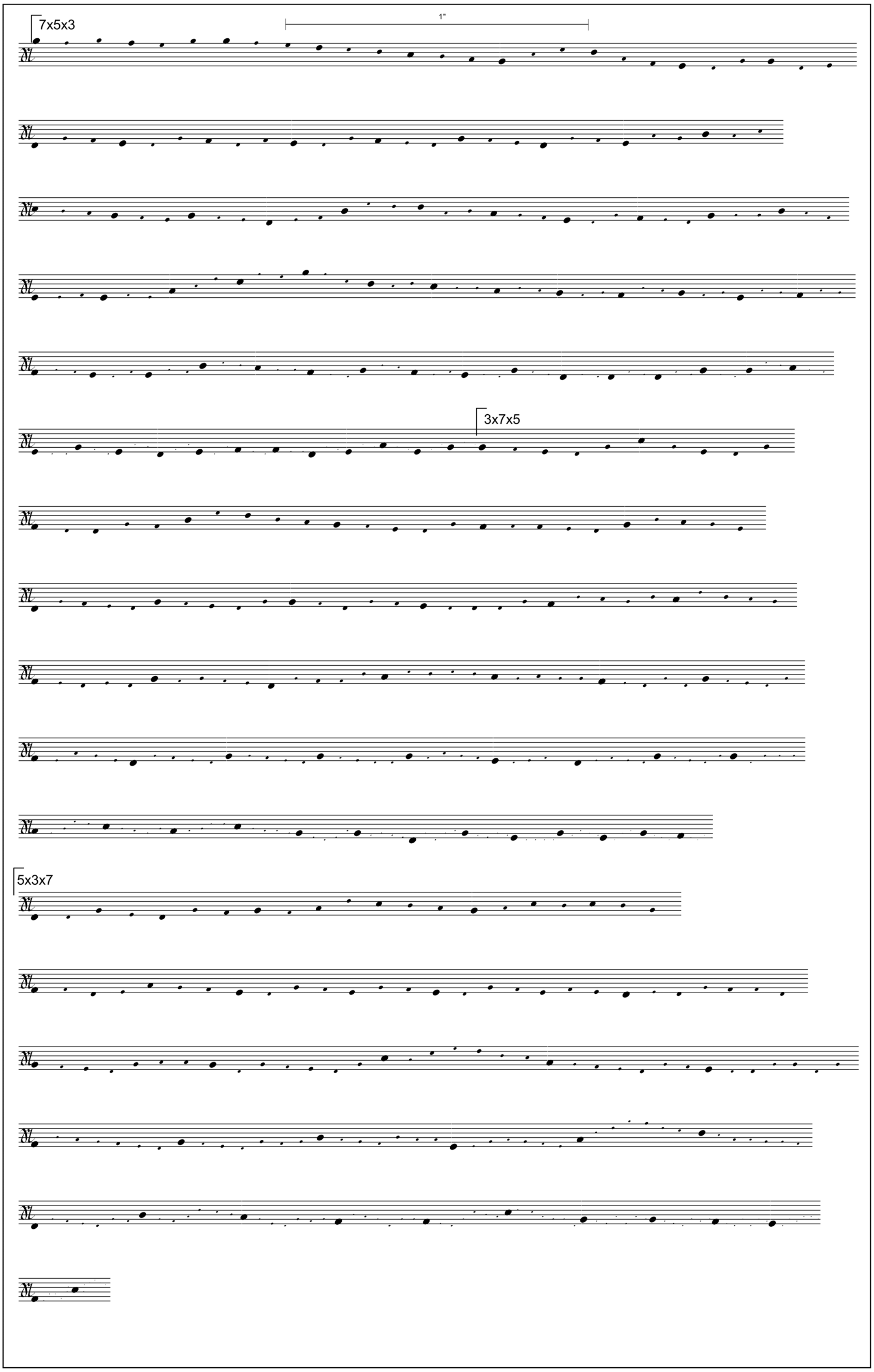


Figure 7. Notation of the Triple Moebius-Band illusion created by Hajdu. The size of a notehead depends on its velocity, which in turn is contingent on its indispensability. Values below a certain threshold will successively fade out while the sequence is sped up. Once the fade-out is complete, a new cycle begins with a rotation of the meter's lowest stratum to the top. The music is in the Bohlen-Pierce scale and notated in its N clef.

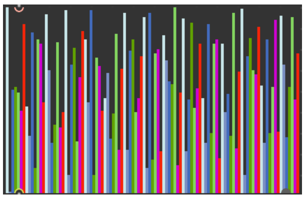


Figure 8. Interpolated profile for meters 3x5x7 and 7x5x3 (fader values [.5, 0, 0, 0, 0, .5]).

Feeding the interpolated profiles into DJster yielded fascinatingly irregular patterns that showed similarity to the rhythms employed by Ligeti in his piano etudes (Figure 6).

2.4 Triple Moebius Band Illusion

Barlow's theory of meter was also validated by an extension of Jean-Claude Risset's illusion of an ever-accelerating rhythm which Hajdu demonstrated as an example for *self-similarity* in music at the 2023 symposium in Hamburg entitled *Ligeti's Labyrinths of Wonderland.*[4] Risset's original illusion consist of fading out (or in) pulses while gradually doubling (or halving) the tempo. In Barlow's terms, it is based on a gradual removal (or addition) of a metric stratum with the stratification divisor of 2. Hajdu modified the illusion by employing meters consisting of three strata with the stratification divisors of 3, 5 and 7. The strata are rotated by gradually fading out the lowest stratum and accelerating the stream of pulses by a factor corresponding to its stratification divisor, and eventually adding the faded-out stratum back to the top.

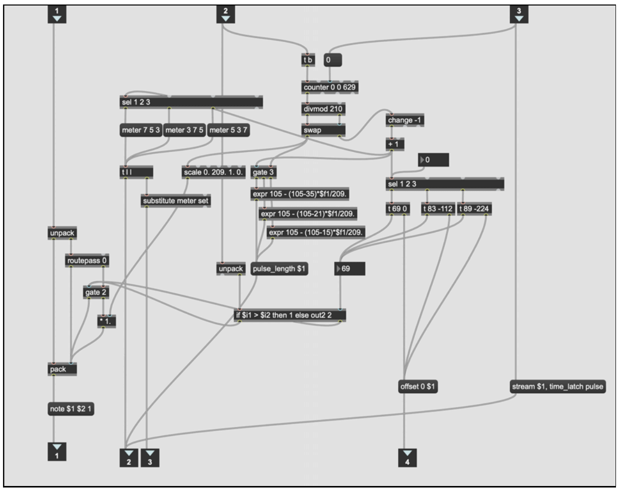


Figure 9. The Max patch driving the Triple Moebius Band illusion. Messages coming out of outlet 1 are sent to a playback device while the messages coming out of outlet 2 control DJster. The second inlet receives indispensability values from DJster.

This can be likened to moving along a Moebius band with three surfaces. The fadeout is correlated to the indispensability values of the metric profiles: The velocities for pulses above a certain indispensability value are kept while the ones equal or below the threshold are being faded out (see Figure 7 for the notation of the illusion which thus also becomes apparent on a visual level). Figure 9 shows how Max communicates with DJster to enable the illusion.

3. PITCH AND MELODY

In *Çoğluotobüsişletmesi* and many of his later pieces, Barlow employed the hierarchical nature of meter and scales to generate music defined by the resulting texture of melodies. To this aim, he introduced the concepts of the *indigestibility* of a number and, closely related, the *harmonicity* of an interval formed by the ratio of two numbers. As the latter term is widely used in psychoacoustics to denote the deviations of partials from an ideal harmonic spectrum, we are using the term *tonal semblance* throughout this paper instead.

3.1 Tonal-Semblance Profiles

We are defining *tonal-semblance* as the inverse of *tonal-distance*: the higher the value, the more similar a tone is to the tonic of a scale[5]. DJster contains preloaded tonal-semblance profiles for 36 scales stored in a Max dictionary. New scales can be added in three ways:

1. As a message to DJster consisting of the message name *add-scale*, a *scale name* and *list of cents values*, the first and the last items forming the replication interval when calculating the tonal-semblance profile (e.g. *add-scale tritones 0 600*). Optionally, a *reference* to four harmonic-energy profiles (narrow, wide, odd-narrow and odd-wide[6]) can be appended to the message. They are used for determining a ratio for a given cent value[7].
2. As an entry in a .cents file containing a *scale name*, a *list of cent values* and, optionally, a *reference* for the harmonic-energy profiles.
3. By dragging a *scala* file from the MaxScore Scala browser (accessible by clicking on the small, encircled B next to "Load .cent file") onto the DJster Scales menu (see Figure 1).

For instance, the cent values for the pentatonic scale with the intervals 0, 200, 400, 700, 900 and 1200 cents are expanded into the following profile ranging from -9600 to 9600 cents, consisting of a *key* and array with *cent value*, *ratio numerator*, *ratio denominator* and *tonal-semblance* value for every pitch:

[4] https://ligeti2023.hfmt-hamburg.de/index.php/toolbar-2-2-2-2-2-2-2-3-2-2-3-2-3-2-2-3-3-2-2-2-2-2/

[5] This similarity has often been explained by how the partials line up between the tones.

[6] The harmonic-energy profiles differ to the extent by which tonally distant intervals are considered or whether even numbers are taken into consideration in the ratios they form, as it is the case with the Bohlen-Pierce and related scales.

[7] If no harmonic-energy profile is given, the profile indicated in the "Harm. E. Profile" menu is used by default.

"pentatonic" : {
"1" : [-9600, 1, 256, -126],
"2" : [-9400, 9, 2048, -62],
…
"36" : [-1200, 1, 2, -1001],
"37" : [-1000, 9, 16, 108],
"38" : [-800, 5, 8, 107],
"39" : [-500, 3, 4, 215],
"40" : [-300, 27, 32, 77],
"41" : [0, 1, 1, 1112],
"42" : [200, 9, 8, 121],
"43" : [400, 5, 4, 120],
"44" : [700, 3, 2, 273],
"45" : [900, 27, 16, 84],
"46" : [1200, 2, 1, 1001],
…
"78" : [8800, 162, 1, 86],
"79" : [9100, 192, 1, 116],
}

Note that this large range is necessary to account for some extreme parameter settings where tonic pitch is in the low range while the pitch window – formed by pitch center and pitch range – is in the high range and vice versa.

Disregarding the large absolute differences between values, the tonal-semblance values (which can be either negative or positive due to the *polarity*[8] of the ratio) are processed to obtain pitch and index pairs and sent to a Max table object with the index determining the likelihood for the associated pitch to occur[9] (Figure 10). The Max *table* object does not actually receive a pitch value, but rather (another) index for an index-to-pitch lookup occurring in the *funbuff* object.

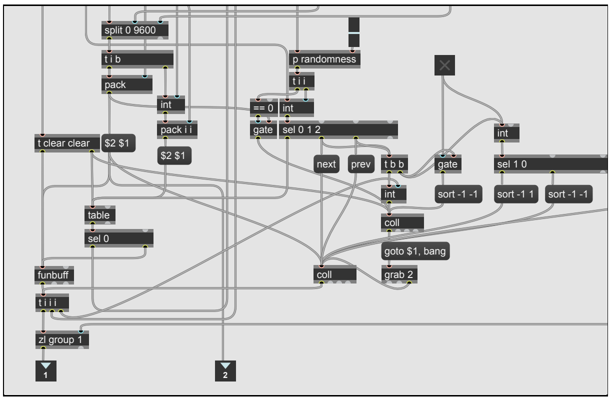


Figure 10. A section from the DJster subpatch implementing the *melodic cohesion* parameter.

3.2 Melodic Cohesion

Autobusk was conceived as a near stateless system, i.e. while doesn't keep a history of past events (e.g. to filter out repetitions), the outcome of each drawing is influenced by the previous one. By introducing the parameter of *melodic cohesion*, we create a continuum between sequential readout and the stochastic drawing of pitches from the tonal semblance profiles, where a value of -100% denotes a descending pitch set, a value of 0% the default probabilistic behavior and +100 an ascending pitch set[10]. This determines, for every event, the likelihood of whether to follow the sequence or draw a pitch stochastically. The pitch sets can be ordered according to the gestalt principles of *proximity* and *similarity*, i.e. to either pitch height or tonal semblance. In the patcher shown in Figure 10, the *randomness* subpatcher sends for every event a number between 0 and 2, their probability controlled by the melodic coherence parameter. These numbers represent the modes of the system (0 = random; 1 = next entry; 2 = previous entry). Note that for each event the complete processed tonal-semblance profile is copied to both *coll* objects and sorted according to pitch height, if the gate, controlled by the *sorted* parameter, is on[11]. If mode is either 1 or 2, the *coll* on the right serves to identify, within the sequence, the index for the pitch of the current event (fed by the red patch chord and stored in the *int* object) in the *coll* underneath, which in turn sends a new pitch to be played. This is how DJster ascertains that a profile be played sequentially despite intermittent random leaps.

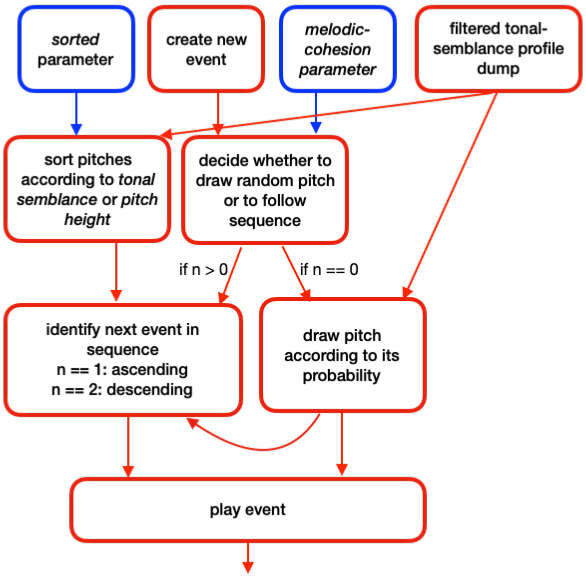


Figure 11. Diagram of the data flow in Figure 10.

3.3 Replacing a Tonal-Semblance Profile

The content of a tonal-semblance profile can be replaced or, when the scale doesn't exist, created by sending a message consisting of the message name *replace-pitchset* and a *scale_name,* as well as a *dictionary* con-

[8] Barlow uses the term to indicate whether the top or the bottom note of an interval is to be considered more stable. His approach convincingly derives the "anti-polarity" of the just fourth or the minor third mathematically but is somewhat arbitrary when applied to tonally distant intervals.

[9] The number of intervals to be considered are affected by the parameters pitch center, pitch range, melody scope, harmoniclarity, and metriclarity values as well as the indispensability for the current pulse.

[10] The transition from a strict sequence to a Markov process of 0th order can be likened to the transition of a solid to a gaseous state where connections between molecules are increasingly broken in favor of free and stochastic movement. This phenomenon has been musically explored by composer Iannis Xenakis from whom the term *stochastic music* originates.

[11] As this is computationally rather expensive, we are hoping to harness the power of the array object recently introduced in Max 8.6.

taining any number of key and value pairs; value being an array consisting of four items *cent value*, *ratio numerator*, *ratio denominator and tonal-semblance value*, such as the following example:

```
{
        "1" : [ 0, "-", "-", 7 ],
        "2" : [ 100, "-", "-", 6],
        "3" : [ 700, "-", "-", 5],
        "4" : [ 600, "-", "-", 4],
        "5" : [ 1100, "-", "-", 3],
        "6" : [ 1200, "-", "-", 2],
        "7" : [ 600, "-", "-", 1]
}
```

Since ratio numerator and ratio denominator are just used for reference and not for calculation, they can be replaced by any placeholder such as “-” or -1. However, the tonal-semblance values ought to be treated as *user-assigned* priority indexes determining the order of the pitches associated with them. With the *Sorted* switch turned off (see Figure 1) and melodic cohesion set to its extremes, motives can be formed which may contain repeated notes.

Note that the tonic pitch and the range parameters (determining the size of the pitch window) should be set to include the motive, whose cent values are relative to tonic pitch (e.g. with C3 as tonic pitch and 6 for pitch range, F#3 would form the center of the window ranging from 0 to 1200 cents). With *Sorted* turned on, the motive would be treated as a scale consisting of 0 100 600 700 1100 1200 cents.

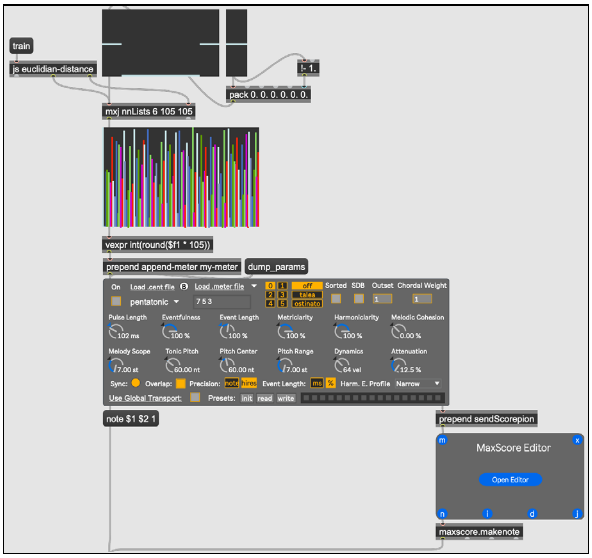


Figure 12. Metric interpolation performed by the nnLists object, a simple MLP object with one hidden layer. Parameter settings—including the stored tonal-semblance and metric profiles—are sent as dictionaries to the DJster Scorepion, a MaxScore Editor plugin via the *dump_params* message.

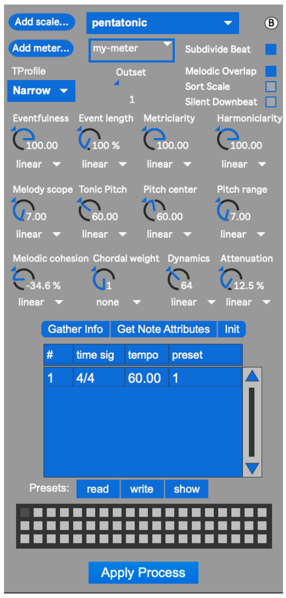


Figure 13. GUI of the DJster Scorepion. Note that the parameter settings have been set according to the sendScorepion message received by the MaxScore Editor.

4. SKETCHING

While DJster was devised as a real-time (synchronous) event generator for Max (and Ableton Live through its Max for Live API), there is also a MaxScore plugin *(Scorepion)* for the asynchronous generation of events. These incarnations can now easily be synced via parameter dumps. When used to compose music, the real-time version of DJster (referred to as DJster RT) can be used for synchronously “auditioning” a stream of events with its parameters set to the desired output. A *dump_params* message causes a dictionary containing the current parameter settings as well as the metric and tonal-semblance profiles to be sent out to a connected MaxScore editor with the DJster *Scorepion* loaded. When prepended by a *sendScorepion* message, the plugin will receive the dictionary and set its buffers and parameters accordingly (Figure 12). In the Scorepion, the user can generate music notation representing the input from DJster RT. In [9] we have presented an earlier version of the Scorepion which is now compatible with the parameter names and ranges of the current version of DJster RT (Figure 13). Since the Scorepion uses the tempo/time signature paradigm of MaxScore, it is recommended that DJster RT make use of the Max *Global Transport* which also uses tempo and time signature rather than absolute time. The pulse length and pulse number are automatically set by the current state of the transport, and the strata of the meter are treated as *subdivisions* of the beat. However, if Global Transport is not being used, *Subdivide Beat* will automatically be switched off in the Scorepion (see Figure 13). This is how the music in Figure 6 was generated.

After performing a score dump, the resulting score can, for instance, be stored in a maxscore.icanvas abstraction, another MaxScore Editor or a Max dict object.

The music can then be used in an arrangement or collage of the material generated: Instead of slavishly following the output of the music generator, the composition thus turns into an interactive and iterative process of auditioning and editing, emphasizing the agency of the composer who thus remains in the "driver's seat."

5. ABLETON LIVE IMPLEMENTATION

The 2023 version of DJster was also turned into an Ableton Live device with Push support. It consists of a wrapper around the Max abstraction with two significant adaptations which we shall describe below.

5.1 Subdivisions

Since Live follows the paradigm of measures, beats and tempo rather than absolute time, the *event length* GUI element is disabled by default, and the value is calculated from the current Live settings. Therefore, a meter chosen in the meter menu is treated as a *subdivision* of the beat rather than defining the entire meter. Subdivisions can be changed within measures: DJster will wait to the next beat and calculate the corresponding pulse number by taking the eclipsed time since the beginning of the measure into consideration (Figure 14). This allows several instances of DJster to be in sync with each other despite switching subdivisions in the interim.

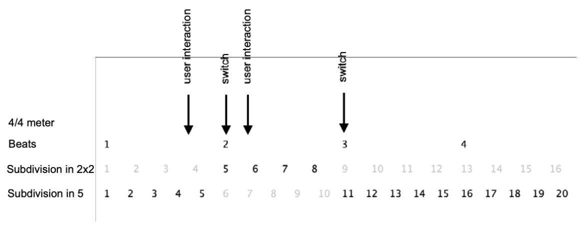


Figure 14. The subdivision of a beat can be changed repeatedly within a measure affecting the pulse number, fundamental for the calculation of indispensability values.

6. OUTLOOK

6.1 Application in Medical Environments: Healing Soundscapes

In 2014, a team from Hamburg consisting of music therapists, music psychologists, medical doctors and composers founded a project called *Healing Soundscapes* with the aim to install and explore the effects of musical soundscape interventions[12] in hospital waiting areas [10]. From the onset, DJster (in conjunction with the MaxScore Sampler) was chosen as an engine to realize the immersive soundscapes, requiring a minimum of four speakers placed in the corners of the (waiting) room. A considerable number of compositions consisting of parameter presets for five instances of DJster and select sound banks have already been created by graduate students of the HfMT.

In 2023, the Healing Soundscapes project became an integral part of the newly founded *ligeti center*, a transdisciplinary joint venture between HfMT, the University of Applied Sciences, the Technical University and the UKE [11]. The project consists of four strands with the following objectives:

1. Establishment of an audio network at the University Heart Center with a central computer delivering streams of generative music created by multiple instances of DJster running on a server.
2. Development of a hardware standalone version of DJster.
3. Development of an intelligent, portable speaker system with each speaker featuring a built-in single-board computer running with the ability to exchange data with the other speakers.

6.2 AI for Analysis and Generation

Current Deep Learning (DL) networks, and Large Language Models (LLM) in particular, are powerful tools which have added an important facet to computer music since about 2016 [12] [13]. One might wonder how a probabilistic music generator might fare in the future considering the current developments of LLMs such as Google's MusicLM [5]. A few experiments with ChatGPT 3.5 prompted to compose a melody for the Bohlen-Pierce scale have shown that it is difficult to achieve sensible results with such general-purpose AI systems and it remains to be seen how soon we will see models capable of capturing cognitive and music-theoretical principles from the music presented to them, to the extent that they can extrapolate from their internal representations of a music corpus and not just create derivative music. In contrast, DJster is not particularly suited to recreating music in specific historical or ethnic styles, but, with its built-in rules grounded in cognition and mathematics, it permits one to venture into the uncharted territories of microtonal music. Considering the strengths of both approaches, the near future might therefore favor a combination of connectionist and stochastic approaches for music generation.

7. CONCLUSION

In this paper, we have shown how a nearly 40-year-old piece of software can be put to good use by embedding it in modern authoring environments and the context of current machine-learning approaches. Its successful application in various use cases validates the approach Barlow took at the time of its inception and the effort undertaken by Hajdu going into further developing this software.

In the introduction, we have outlined the possibility of establishing a recursive feedback loop consisting of the analysis and synthesis of musical material. The idea behind this is not fueled by the desire to reproduce an input as faithfully as possible but rather by a process of *resonance* where a generator responds to a musical

[12] https://tinyurl.com/5c753mrc

input according to its own affordances. This could also be seen as a metaphor for intercultural or interspecies dialog. This approach could be expanded by using text prompts describing the preferred character of the music or an emotional state, very much like the way Dall-E, ChatGPT or MusicLM create content based on a given text input.

When pondering the question of using AI for the creation of music, it dawns on us that it may not be as straightforward as it might seem. Technological progress doesn't follow an exponential curve, but rather a sigmoid curve. This implies that progress can be made fast, but perfection takes a long time to achieve. Since in art and science, its practitioners strive towards perfection, we postulate that it may take an extended amount of time to train a DL network to achieve comparable depth and perfection that we have grown to admire in the art of the great masters.

Acknowledgments

We would like to thank Lin Chen, James Tsz-Him Cheung and Alessandro Anatrini for their contributions to this paper as well as acknowledge the Federal Ministry of Education and Research in Germany (BMBF) and the Hamburg Authorities for Science, Research, Equality and Districts (BWFGB) for their support of this research.

Examples

Supporting materials are available at:
https://tinyurl.com/bdkc2ass.

8. REFERENCES

[1] C. Barlow's AUTOBUSK user manual "AUTOBUSK: A real-time pitch and rhythm generator" for the Atari ST version is available as report no. 44 (2001) in the series "Musikinformatik & Medientechnik".

[2] G. Hajdu. "Resurrecting a Dinosaur – the adaptation of Clarence Barlow's legacy software AUTOBUSK," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'16,* 2016, pp. 181–186.

[3] X. Fu, *Kommunikation und Technologie in dem Tanztheaterstück enactment::interchange*. Hofheim: Wolke, 2023.

[4] G. Hajdu, B. Carey, G. Lazarevic, and E. Weymann, "From atmosphere to intervention: the circular dynamic of installations in hospital waiting areas," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), 2017*, pp. 364–369. https://doi.org/10.5281/zenodo.1176284

[5] A. Agostinelli et al., "Musiclm: Generating music from text," *arXiv* preprint arXiv:2301.11325, 2023.

[6] S. A. Taylor, "Hemiola, maximal evenness, and metric ambiguity in late Ligeti," *Contemporary Music Review* 31, no. 2–3, 2012, pp. 203–220.

[7] C. Barlow, "Two Essays on Theory," *Computer Music Journal,* vol. 11, no. 1, 1987, pp. 44–60.

[8] C. Barlow, "Bus journey to parametron," *Feedback Papers*, vol. 21–23, 1980.

[9] B. D. Smith and G. E. Garnett, "Unsupervised Play: Machine Learning Toolkit for Max," *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2012. https://doi.org/10.5281/zenodo.1178419

[10] E. Weymann, "Healing Soundscape," *Lexikon Musiktherapie* 3, 2020, pp. 239–243.

[11] G. Hajdu et al., "The ligeti center: Ligeti and his vision of a center for computer music in Hamburg," *Proceedings of the International Computer Music Conference*, Shenzhen, 2023, pp. 81–85.

[12] A. van den Oord et al., "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[13] A. Roberts et al., "Magenta studio: Augmenting creativity with deep learning in ableton live.", 2019. [Online]. Available: http://research.google/pubs/pub48280

SOUND SYNTHESIS NOTATION APPLIED TO PERFORMANCE: TWO CASE STUDIES.

Pierre-Luc Lecours
Université de Montréal
CIRMMT student member
pierre-Luc.lecours@umontreal.ca

Nicolas Bernier
Université de Montréal
CIRMMT regular member
nicolas.bernier@umontreal.ca

ABSTRACT

This article investigates the specificities of music writing and interpretation on the modular synthesizer. Based on two musical notation experiments, it will discuss the issues first from the point of view of the composer and then from that of the performer. This article will begin by presenting the notation approach in the composition of Pierre-Luc Lecours's piece *Poussière de soleil* (2022) performed by Ensemble d'oscillateurs. Then it will analyze the stages involved in creating an interpretation of Nicolas Bernier's composition *Transfer for 10 monophonic synthesizers* (2022).

These two experiences revealed issues and strategies used when writing and interpreting a piece with modular synthesizers, pointing toward a notation framework for this instrument.

1. INTRODUCTION

While performance on modular synthesizers is common in contemporary electronic music, the notation conventions of complex parameters (such as timbres or modulations) for this form of performance are still rare. After more than 60 years since the advent of the synthesizer, transmitting musical indications for this kind of instrument is still problematic. Commonly, artists who perform on modular synthesizers also act as composers (Suzanne Ciani, Caterina Barbieri, Alessandro Cortini to name a few). Their notation (if any) is commonly intended for personal use, not for communicating the piece to other performers.

In this article, we define the modular sound synthesis performer as a specialist in the interpretation of a repertoire linked to any electronic or network of electronic instruments producing synthesized sounds which must be programmed prior to their use and whose parameters (pitch, rhythm, and timbre) can be controlled manually or automated with voltage and/or digital control automations. These synthesis instruments can produce sounds from both analog and digital circuits, excluding sounds from conventional acoustic instruments. This definition therefore also excludes electronic instruments whose synthesis systems are pre-programmed (presets-based) and whose interpretation parameters are mainly pitches and their intensities (e.g. Yamaha's DX7 or SY-77 synthesizers). This definition would typically involve building a synthesis system to meet the interpretation needs of the piece to be played.

This article looks at some of the specific features in writing for modular sound synthesis instruments, examining the issues first from the point of view of the composer and then from that of the performer. These two positions are illustrated by two experiments in the notation for the performance of modular sound synthesis.

The first case explores the challenges of notation for modular synthesis from a composer's perspective. This section emphasizes the strategies utilized in the notation of sound synthesis systems. These strategies encompass a combination of conceptual and systemic approaches, the merits and drawbacks of which will be evaluated. Additionally, we will examine how the communication of ideas and musical actions to be executed by performers also involves a blend of descriptive and prescriptive methods.

The second case is the first part of a study in which we sought to determine the learning stages and strategies used in the creation of an interpretation of a piece written for analog synthesizers. The aim is to begin to establish points of convergence and divergence between modular synthesizer performance and more traditional instrumental practice, in order to highlight certain specificities.

These two experiments highlight specific features to be taken into account when writing music for the modular sound synthesis, and these will be presented in the discussion section.

2. CASE 1: NOTATION APPROACH FOR MODULAR SYNTHESIS ENSEMBLE IN PIERRE-LUC LECOURS COMPOSITION *POUSSIÈRE DE SOLEIL*

The piece *Poussière de soleil* was composed in the autumn of 2022 for Ensemble d'oscillateurs. Directed by Nicolas Bernier, the ensemble consists of ten musicians playing on different electronic instruments, 10 Moog Mother-32 synthesizers in this case. The ensemble developed its own way of writing and reading scores which are commonly made of large graphics scrolling from right to left on a screen.[1]

The ensemble created a vectorial graphic model that can be adapted according to the composer's compositional needs while standardizing the transmission of the three main performative parameters of the oscillators: frequencies, amplitudes, and human-made modulations [1].

The ensemble notation was initially specifically created for sine wave interpretations as the ensemble's first instruments were sine wave generators. The scores are shown on small screens placed in front of each musician (Fig. 1). A first attempt at more complex timbre notation has been made for the ensemble with Nicolas Bernier's composition *Transfer for 10 monophonic synthesizers* using a hybrid approach between prescriptive abstraction (prescriptive indication referring to a hypothetical synthesizer) and instrument-specific notations (the Moog Mother-32 in this case).

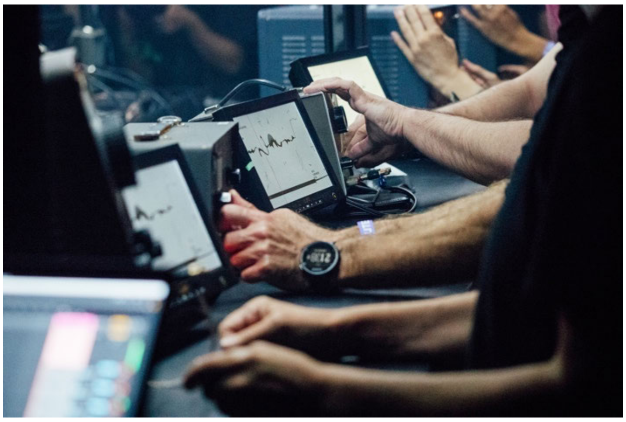


Figure 1. Ensemble d'oscillateurs with their screens and sine wave generators.

While starting from the ensemble notation methodology for the pitches, durations, and intensity, Lecours in his composition *Poussière de soleil* choose to notate the synthesis system and its manipulation with a systemic conceptual abstraction approach. This method implies the notation of the synthesizer patch and its manipulations to be not specific to a particular instrument configuration. Instead, it referred to a conceptual modular synthesis system, meant to be reinterpreted by performers using their instruments.

When composing his piece, Lecours intentionally avoided studying the specifics of the Mother-32 synthesizer. His goal was to write the playing instructions that could be interpreted on any modular or semi-modular synthesis system. These systems needed to include at least an oscillator (with a choice of basic waveform), a noise generator, a filter, an envelope generator, and a chromatic controller. The composition process occurred in two stages, revealing several distinctive features in the approach.

2.1. Composition and Notation System

To compose his piece, Lecours created most of the musical material with his own modular synthesizer. He designed a synthesis system with which he recorded different ideas. These recordings were then segmented into ten voices, which he transcribed onto a score[2] (Fig. 2). The segmentation was based on the three modular synthesis systems distributed over the 10 voices as follows: First system: synthesizers voices 1 to 3, second system: synthesizers voices 4 to 7, third system: synthesizers voices 8 to 10

This division was designed to allow sectional playing by the ensemble, with the performers playing similar musical material to the others in their groups.

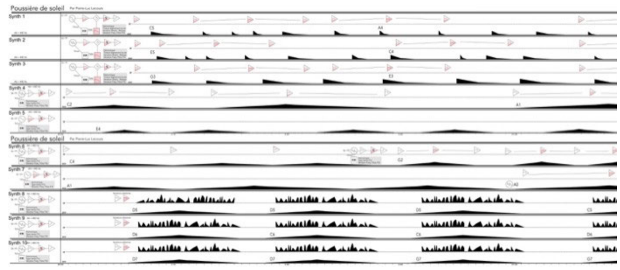


Figure 2. Score of the first minute of *Poussière de soleil*.

There are three types of indication on the score. The first is located in the lower part (Fig. 3) of each voice and represents: pitch (indicated by a letter plus its octave), duration (indicated by the black shape horizontally), and intensity (indicated by the black shape vertically from ppp to fff).

Figure 3. The first type of indication (pitch, duration, intensity).

The second type of indication is located in the upper part of each voice (Fig. 4) and indicates the manipulations to be performed (e.g. sound filtering) or general indications like acceleration or deceleration (Fig. 5). The activation of the parameter to be manipulated is indicated by the amount of *red* present in the symbol; in the example in Figure 4, the low-pass filter will be almost closed at the start of the manipulation, opening to 2/3 at the moment when the sound intensity is at its highest, and finally returning to the initial state at the end of the note.

[1] Several ensembles use this approach like the Decibel New Music Ensemble with the Scorereader software they develop (Decibel Score Player).

[2] The scrolling score along with the recording can be accessed on Vimeo: synthesizer voices 1–5 (https://vimeo.com/923558953/527aaed017) and synthesizer voices 6–10 (https://vimeo.com/923569353/75e9775256).

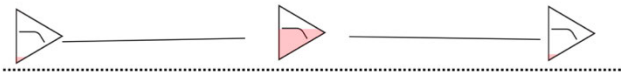


Figure 4. The second type of indication (parameters manipulations).

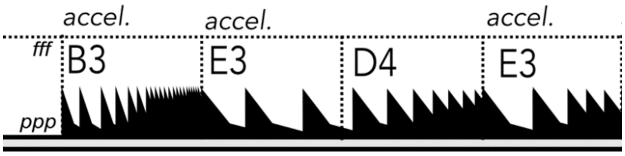


Figure 5. The second type of indication (general indications).

The third type of indication represents the sound synthesis systems to be created (Fig. 6). These indications are designed to convey maximum information using a minimal number of symbols. The parameters to be manipulated are indicated in the text box. The symbol system used is the one proposed by the book Patch and Tweak [2]. In Figure 6, the performer should be able to play chromatic notes on a controller that will give the pitch information to an oscillator outputting a square waveform. The square wave signal then passes through a low-pass filter whose output is modulated by an envelope triggered by a gate whose attack and resonance must be adjustable. The modulated signal then passes through a reverb effect activated at 50% and finally goes to the sound console. The graphical indications are supplemented by text that adds details, as in the example “Timbre légèrement bruité” (Slightly noisy timbre), which can be interpreted more freely by the performer.

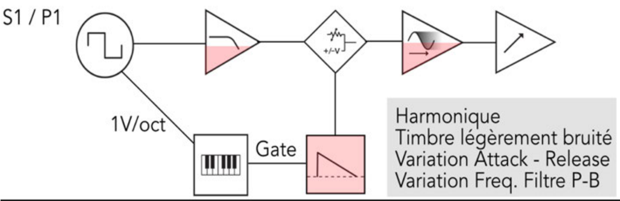


Figure 6. The third type of indication represents the sound synthesis systems to be created.

This approach seeks to represent the most important elements of the system for interpretation in order to optimize the readability in a performance context and its realization on different synthesis systems (we will discuss these issues in the discussion section). The performer then has to find idiomatic solutions for their instrument to respond to the score’s indications. It is interesting to note that during the rehearsals with the ensemble, the musicians generally devised different playing systems based on the same indications, with very similar sound results, but with enough timbral differences that it did enrich the sound image of the piece.

2.2. Working With the Ensemble

Following the ensemble’s first rehearsal, Lecours finalized the composition of the piece. He attended most of the rehearsals with the ensemble and its conductor, which enabled him to imagine new ways of bringing the musical material into dialogue with the playing of the performers and ultimately that of the conductor. For example, in the last section of the piece, where this dialog led to musical material in which the scrolling score format did not allow the notation of rhythmic metrical material with synchronized acceleration. It resulted in the conductor beating time and directing the acceleration as well as indicating the crescendos and decrescendos.

This back-and-forth between the score and the ensemble’s performance enabled Lecours to make necessary adjustments and corrections in the notation. It was interesting to observe the performers’ relative independence concerning the conception of their sounds; they found their own solutions on their instruments to interpret their respective parts.

3. CASE 2: SOUND SYNTHESIS (RE)PERFORMANCE APPROACH OF NICOLAS BERNIER’S *TRANSFER FOR 10 MONOPHONIC SYNTHESIZERS.*

The first case study discussed in the previous section presented notation strategies used in a composition by Pierre-Luc Lecours. In this section, Pierre-Luc Lecours is acting as a performer, interpreting a piece Nicolas Bernier originally wrote for 10 synthesizers.

A few elements must be specified about the choice of this piece before discussing the performance issues and strategies. The score was not written *specifically* for the modular synthesizer. Indeed, the piece required translating indications originally intended for the Mother-32 synthesizer into the modular synthesis instrument used for this reperformance. Ideally, the study would have been based on a composition specifically tailored for the modular synthesizer. Although such compositions do exist, they are rarely notated in a score format. When they are, these compositions are often written for a specific modular sound synthesis system, making it difficult to perform without a specific combination of modules. Another limitation was that Bernier’s piece was composed for ten performers and thus intended to be played by an ensemble. While this characteristic did not hinder the study’s objectives, the musical material was not crafted for solo performance, thus influencing Lecours’s understanding of the piece’s nuances and development.

3.1. Methodology

Our methodological approach to modular synthesizer performance was largely based Héroux and Fortier’s [3] study of the interpretation creation process while also borrowing from Chaffin et al.'s [4] study of the different stages involved in the learning of a new piece.

Elements from grounded theory [5] were used to analyze the interpretation process: based on interviews and infused by the notions of ‘explicitation’ interview [6] with the composer and self-confrontation [7]. From grounded theory, we retained the creation of concepts (in our case a new stage in the learning process defined by Chaffin and the double reading of the score) in response to the analysis of the data obtained. Two approaches were adopted to obtain the data, and these are based on the experiment presented in Héroux and Fortier [3].

Firstly, Lecours filmed himself rehearsing the first synthesizer part, commenting on his actions and reflections aloud. Reviewing these recordings allowed him to comprehend his approach at various stages of the interpretation creation of the first voice. Secondly, Lecours maintained a practice diary for all the rehearsals, describing in detail the actions performed, the progress of the work, and the relevant observations made during the rehearsals. This approach facilitated monitoring his progress and adjusting his practice methods accordingly.

Once satisfactory interpretations of the different parts were achieved, Lecours recorded the ten voices of the piece in a studio setting. This recorded version was then shared with the composer for feedback on the interpretation. In parallel, weekly interviews with composer Nicolas Bernier took place, where discussions about the work conducted during the practice sessions occurred. These interactions provided valuable insights and feedback on the progress made during the rehearsal process.

3.2. The Interpretation Process

The process of creating this interpretation spanned four work sessions, each lasting approximately three hours. In these rehearsals, the ten voices were learned sequentially, rehearsed, and then recorded individually. Like in the previous case, the score was also based on the ensemble notation convention (Fig. 7)[3] where the upper half of the staff of each instrument shows pitches and other indications and the lower half of the staff shows the envelope and amplitude (Fig. 8).

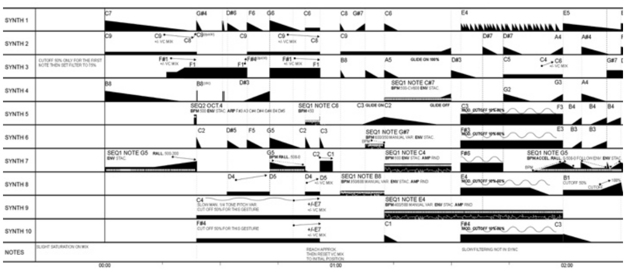


Figure 7. The beginning of the *Transfer for 10 monophonic synthesizers* score.

Figure 8. An isolated voice with indications of notes and intensities.

For this piece, Lecours opted for an instrument configuration that allowed him to work with most of his synthesis modules. He used the two-octave keyboard of his Arturia Minibrute 2 synthesizer. It should be noted that the modular synthesizer was not pre-configured; it had to be organized into a system with cables that met the requirements of musical interpretation with each new voice. The configuration of modules used allowed great versatility in terms of interpretation and possible aesthetics, enabling him to interpret the indications in the score.

3.3. Organization of Work Sessions

Lecours began the first session by doing an overview of the score to understand the types of sounds and controls he needed to implement for the performance. This stage aligns with Chaffin's "scouting-it-out" phase [4]. While reviewing the video recording of the session, his comments revolved around two main aspects: the specific actions to be executed (such as notes, octave changes, rhythmic patterns, etc.) and how to configure his modular system to respond to these musical instructions.

At this point, there was a dual interpretation of the score: one from the performer's perspective and another from the sonic system designer's viewpoint. The sonic system designer's reading is one in which the performer thinks about how to organize the instrument's synthesis system to execute the musical indications expressively on the one hand, but also in such a way that it is possible to execute the different indications through the piece without radically reconfiguring the synthesizer.

With modular synthesizers, generally analog, the configuration of the synthesizer requires a balance between the search for timbre and the expressiveness of the sound, with the resources available for setting up the different systems needed to execute a part. These considerations stem from the limitations of each modular instrument in terms of connections and modules.

After completing the initial overview, Lecours proceeded to study the composer's instructions on programming sounds for the Mother-32 synthesizers. He used these guidelines to configure the sound synthesis system on his instrument. Lecours had to adapt the instructions provided by Bernier to apply them to his synthesizer. As some instructions were specific to the Mother-32, he modified certain control approaches and disregarded instructions that were too specific to convert for his instrument. Nonetheless, Lecours managed to establish a fundamental playing system, allowing for an initial reading of the piece.

This stage of the process was akin to what Chaffin [4] refers to as "section-by-section." Lecours focused on the first synthesizer part, breaking it down into short sections played in loops. This approach highlighted the piece's structure, providing him with a deeper understanding of its different sections and allowing him to pinpoint specific technical challenges.

At this stage, there was also a dual reading of the score, with the performer on one side and the system sound designer on the other. This dual reading created a feedback

[3] The scrolling score along with the recording can be accessed on Vimeo: https://vimeo.com/923571511/8941f2ae15.

loop between the performance work and the system sound design where, during the rehearsal, he would perform a section, optimize the synthesis system, then perform again, move on to the next one, make changes to the connections, adjust the parameters, etc. It was also at this stage that he was able to identify some of the technical issues and, that he began to take liberties in the sound system conception to make certain sounds more interesting and expressive.

In the second work session, Lecours began with a swift overview and focused on short sections of the second synthesizer part. He then revisited the first part, refining transitions and performing a complete run-through before proceeding to record the performance. These final stages of work on the first voice align with what Chaffin et al. refer to as the "gray stages" and "maintenance." The dual effort of interpretation and sound design persisted until the recording phase.

During the third and fourth working sessions, the same approach was applied to parts three through ten, and no significant issues were encountered. It should be noted that as Lecours advanced in the process of learning and recording the various voices of the piece, his understanding of the ideas and functions of each section deepened. Consequently, the artistic appropriation, as mentioned in Héroux and Fortier's 2014 article [3], developed relatively late in the process, which may be attributed to the nature of the piece, which is originally for ten performers.

3.4. Variation on Stages of Chaffin et al. (2003)

One of the most interesting aspects of this research was to observe the similarities and differences in relation to the four-stage model of instrument work during the creation of an interpretation proposed by Chaffin et al. [4]: 1 – scouting-it-out, 2 – section-by-section, 3 – gray stage, 4 – maintenance.

By analyzing the data collected, we can observe, first of all, an additional stage which we call "creation of the sound synthesis system." This stage involves structuring the performance system of the modular synthesizer to execute the specified musical instructions and produce sounds in alignment with the timbral requirements outlined in the score. In this study, this stage was positioned between the first and second phases of the process.

Another aspect that came to our attention in analyzing the data from this research was the notion of a double reading of the score in the learning stages, involving a constant oscillation between interpretation and system sound design. It appeared that the emphasis of this dual interpretation shifted throughout the process of learning the piece. Initially, there was a significant focus on the role of the sound designer, aligning with the first two stages outlined by Chaffin et al. [4]. However, as the learning process progressed, the focus gradually shifted towards a more pronounced emphasis on interpretation during the last two stages.

The similarities and differences observed can be visualized in Figure 3. It is essential to mention that these results reflect the specific context of the present experiment. To validate this analysis comprehensively, it will be crucial to replicate a similar experiment with multiple subjects.

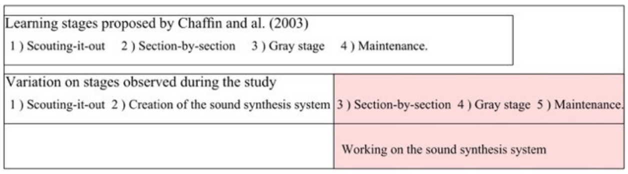


Figure 9. Similarities and differences with the learning stages proposed by Chaffin et al. (2003).

3.5. Specific Features of *Transfer for 10 monophonic synthesizers* Music Notation

This study has highlighted specific characteristics and challenges related to Bernier's piece notation. The first specificity of the score is that it is a scrolling score, which is a format that is suitable for pieces requiring proportional temporal representation [8]. In the case of Bernier's work, the score format was well suited to the musical material, as the piece was not based on a traditional metrical structure, it would probably have been more difficult to read with a traditional writing style.

The second element to be discussed concerns the indications for creating the synthesis and playing system, which here were partly written for Mother-32. Some of Bernier's musical indications had to be abstracted and conceptualized for a system other than the Mother-32. This process of abstracting and conceptualizing Bernier's indications led to liberties being taken in the design of the synthesis systems and to certain sounds being adjusted to suit certain personal aesthetic tastes.

That said, the clarity of the indications in Bernier's piece made for a relatively fluid reading.

3.6. Towards a Definition of the Stages Involved in Creating an Interpretation With a Modular Synthesis System

The results of this research work should be seen as the beginning of a larger research project, that said, certain observations and reflections can be drawn from it.

Firstly, the recorded version was well received by the composer when he listened to it; he was surprised by the match between his intentions and the interpretation. This can be interpreted as a writing success on the part of the composer, but also that this process, i.e. writing a piece for a modular sound synthesis interpreter via a score without accompanying sound support, is possible.

Secondly, there are certain points of convergence and divergence in the stages of creating a musical performance. Based on the stages proposed by Chaffin et al. [4], it is possible to observe that Lecours went through the same process as those mentioned above, except for an additional stage that we have named *Creation of the sound synthesis system*. Additionally, there was a dynamic interaction between sound design and performance system design throughout each phase of performance creation.

There are several aspects of this study that would merit further investigation. One of these is the concept of the *Technical image* that emerged during the process. We define this category as the construction of a coherent representation between the synthesis system put in place and the performance needs of the piece. This idea will require a much more serious analysis of the different strategies and stages involved in the creation of the sound synthesis and playing system. For the moment, the data collection that we carried out during the experiment was not sufficient to fully monitor the process.

4. DISCUSSIONS AROUND WRITING AND INTERPRETING NOTATED SOUND SYNTHESIS PIECE

These two experiments made possible the observation of specific aspects linked to performance and composition for modular sound synthesis instruments. These particularities have led to the use of specific writing strategies, which will be discussed here.

The discussions will revolve around the two levels of notation explored in this article: the one governing the creation of the sound synthesis system (how to configure one's modular sound synthesis system) and the other concerning the sound manipulations to be performed (producing sounds (with or without pitches), timbral and rhythmic variation).

4.1. Creation of the Sound Synthesis System

Interpreting a musical idea with a modular synthesis system requires the design of a connection network on the instrument that allows the required manipulations to be carried out. As we saw in the study of the interpretation of *Transfer for 10 monophonic synthesizers*, the design of a sound synthesis system to respond to the musical indications in the score is a fundamental stage in the process. The notation must therefore take into account several parameters in order to achieve the desired result.

4.1.1.Variability of Modular Sound Synthesis Instruments

One of the many features of this instrument is its versatility. Performers of modular sound synthesis build their instruments by choosing modules or programs that specifically meet their musical needs. Some opt for hybrid analog/digital or modular/semi-modular approaches, leading to virtually endless combinations.

This diversity implies that each performer becomes an expert in their unique system configuration. Due to the absence of standardization (except for electrical control and communication in the case of the Eurorack format and electronic protocols), approaching notation in a highly specific manner becomes challenging without mandating a precise system configuration. Moreover, an overly hardware-specific notation approach is not sustainable, given that synthesizer and digital program modules have a limited lifespan. It also appears incompatible to us with the culture of modular sound synthesis performance.

4.1.2. Modular Sound Synthesis Performance Culture

A second particularity of this instrument comes from its performance culture. We consider that there is a form of expressiveness specific to the performance of a modular synthesizer that comes from the design of the system itself. The personalization of each modular synthesis system is often the result of a search for possibilities of sound generation and modification that are specific to each musician. This personalization creates a sound identity specific to each performer. Moreover, performers develop a set of strategies to craft sound systems that align with their aesthetic preferences and musical ideas, utilizing the available modules and connections. This set of strategies thus becomes highly personal to the performer and his system.

Given this distinctiveness, we propose that creative expression through the construction of a modular sound synthesis system should be taken into account when writing the score.

4.1.3.Notation of a Sound Synthesis System: Opting for a Systemic Abstraction

The variability of modular sound synthesis systems and the intricacies involved in their construction led us to approach this type of notation by seeking to abstract synthesis concepts rather than represent them in their entirety.

This abstraction takes the form of a network of interconnected icons designed to represent the structural logic of the system. These icons highlight the fundamental units of the system's construction and the parameters that need manipulation (in red in the score of *Poussière de soleil* in Fig. 8, for example). Additionally, textual indications complement this iconographic representation, providing performers with interpretative freedom. This systemic abstraction approach allows the musicians to build a system using different tools from those used by the composer, and to choose how they will carry out the manipulations indicated.

For instance, a performer instructed to filter a sound (as shown in Fig. 6) might opt to use their own CV controller to manipulate the cutoff frequency of the low-pass filter instead of the module's designated knob, especially if this choice accommodates other necessary interventions. While such specific directions may not be explicitly mentioned in the score, this more abstract form of notation provides freedom and flexibility in system design that, we believe, aligns with the unique characteristics of modular sound synthesis performance outlined earlier.

4.2. Sound Manipulation Notation

The instructions for sound manipulations, including sound production (with or without pitches) and timbral and rhythmic variations, are intricately tied to the modular sound synthesis system producing them.

In the notation experiment for *Poussière de soleil* presented in this article, two types of indications are utilized: one for pitches, duration, and intensity, and another for specific parameter manipulations (such as filtering, LFO oscillation speed, etc.).

In the first case, the approach used is the one generally employed by the Ensemble d'oscillateurs. However, for this piece, a more conventional method of notating pitches, durations, and intensities could have been considered.

In the second case, the instructions for manipulating precise parameters pertain to specific aspects of the performer's synthesis system setup. A symbol from the sound synthesis system construction guide (depicted in Fig. 8) is used to denote the manipulation to be performed. This approach allows the performer to make a correlation between his system and what he has associated with a given symbol and the manipulation he has to perform.

4.3. Sound Manipulation and Systemic Patch Notation, Between Prescriptive, Descriptive, and Conceptual Approach

These specific features of the instrument and its performance make it difficult to adopt an exclusively prescriptive or descriptive approach to notation, more traditionally observed in contemporary notation. In the piece *Poussière de Soleil* presented in this article, the way the instructions are notated falls somewhere between the descriptive, prescriptive [9], and conceptual model approaches.

The descriptive approach was employed to indicate pitches, durations, and intensities in the notation of *Poussière de Soleil*. This strategy proved to be effective and unambiguous during rehearsals of the piece and is similar to the traditional ways of representing these parameters.

Then, for the notation of sound synthesis systems, a conceptual approach representing the relations between entities was used. The conceptual approach means that we abstract a specific sound synthesis system in order to extract its fundamental logic and represent it in the form of global diagrams that allow the relationships between the different elements to be understood quickly. This strategy was chosen because it simplifies comprehension in a reading situation, the representation is clearer, and it allows the represented concepts to be applicable to various modular sound synthesis systems.

In the notation for manipulating precise parameters (Fig. 6), the approach used falls somewhere between conceptual and prescriptive representation. In this figure, the instructions are generally prescriptive (e.g., "vary the cutoff frequency of a low-pass filter") but refer to a conceptual representation of the previously constructed sound synthesis system.

For instance, considering the sound filtering indication, if the approach had been purely prescriptive, it would require specifying the manipulation of a potentiometer on a specific system.

To address the variability inherent in modular sound synthesis systems and the idiosyncratic approaches of their performers, we propose that indications for manipulating specific parameters should be prescriptive while referencing the conceptual representation of the system. This hybrid approach accommodates the unique features of modular setups and provides clarity to performers.

4.4. Limitations of the Proposed Strategies

It is important to acknowledge certain limitations to the strategies proposed here. The two pieces whose notations have been presented propose relatively simple sound synthesis systems. It would be valuable to test these proposals in the composition of pieces where synthesis and timbre work are more complex.

Additionally, *Poussière de Soleil* was played in December 2022 and November 2023, the composer was present at the rehearsals and directed the first performance. This involvement might have facilitated the transmission and execution of the sound synthesis systems and the performance of the piece.

5. CONCLUSIONS

Notation for modular sound synthesis remains a relatively underexplored topic, and its interpretation is still an emerging practice. This article has identified important specificities related to composition and the interpretation for and with the modular synthesizer, as well as unique challenges associated with it. The proposals derived from the two experiments presented were explained and contextualized.

Several follow-ups to this research project are planned. Firstly, a series of studies for solo performers or small ensembles is currently in development and will be presented in the upcoming year. Secondly, the strategies outlined in this article will be tested in a composition featuring much more complex synthesis systems. This will enable the examination of the limitations of the proposals outlined in this article and provide valuable insights into the evolving field of modular sound synthesis notation and interpretation.

Acknowledgments

We would like to thank the performers of the Ensemble d'oscillateurs for their work on the piece. The Social Sciences and Humanities Research Council of Canada (SSHRC) for funding this research, the Laboratoire formes • ondes (LFO) de l'Université de Montréal and the Centre for Interdisciplinary Research in Music, Media, and Technology (CIRMMT).

6. REFERENCES

[1] N. Bernier, "Observations on Performing Sine Waves with an Oscillator Ensemble," *Leonardo* 55(2), pp. 161–165, 2022.

[2] K. Bjørn and C. Meyer, *Patch & Tweak. Exploring Modular Synthesis*, Bjooks Media, 2018.

[3] I. Héroux and M. S. Fortier, "Expérimentation d'une nouvelle méthodologie pour expliciter le processus de création d'une interprétation musicale," *Les cahiers de la Société québécoise de recherche en musique* 15(1), pp. 67–79, 2014.

[4] R. Chaffin, G. Imreh, A. F. Lemieux, and C. Chen, "Seeing the Big Picture: Piano Practice as Expert Problem Solving," *Music Perception* 20(4), pp. 465–490, 2003.

[5] P. Y. Martin, and B. A. Turner, "Grounded Theory and Organizational Research," *The Journal of Applied Behavioral Science* 22(2), pp. 141–157, 1986.

[6] P. Vermersch, *L'entretien d'explicitation,* Paris: ESF, 2006.

[7] J. Theureau, "Les entretiens d'autoconfrontation et de remise en situation par les traces matérielles et le programme de recherche 'cours d'action' ", *Revue d'anthropologie des connaissances* 4(2), pp. 287–322, 2010.

[8] C. Hope and L. Vickery, *Screen scores: New Media Music Manuscripts*, p. 226, 2011.

[9] C. Seeger, "Prescriptive and Descriptive Music-Writing," *The Musical Quarterly* 44(2), pp. 184–195, 1958.

EMA: AN ANALYTICAL FRAMEWORK FOR THE IDENTIFICATION OF GAME ELEMENTS IN GAMIFIED SCREEN-SCORE WORKS

Takuto Fukuda
CIRMMT, Concordia University
takuto.fukuda@mail.concordia.ca

Paul Turowski
ICCaT, The University of Liverpool
paul.turowski@liverpool.ac.uk

ABSTRACT

Gamified compositions – music involving game elements (e.g., avatars and life points) – have been booming in the field of interactive computer music. However, only a few studies have addressed which game elements engender the sense of playfulness in performer-computer interactions in music. This gap may exist because existing analytical frameworks primarily focus on identifying game elements in consumer products rather than musical compositions. To address the lack of analytical frameworks for gamified musical works, this paper proposes the Expanded Motivational Affordances (EMA) model as an analytical framework for identifying game elements in gamified screen-score works. Through an analysis of *Super Colliders* by T. Fukuda and *SQ2* by P. Turowski as case studies, this paper provides a comprehensive list of game elements and discusses what motivational needs for performers these elements satisfy. The EMA model with the resulting list of game elements aims to assist composers in gaining a better understanding of performer-computer interactions in gamified screen-score works. It enables composers to analyze and design such interactions more effectively in their future compositions, enhancing the overall experience for performers.

1. INTRODUCTION

In recent years, the application of gamification, defined as "the use of game design elements in non-game contexts" [1], has gained significant traction in the realm of interactive computer music [1]. The term "gamified" has emerged to describe a specific type of composition that integrates game elements–building blocks that support game mechanics–such as avatars and life points as exemplified by "gamified audiovisual works" [3] and "gamified screen-score" [4]. Noteworthy pieces in this vein include P. Turowski's *Hyperions* (2014), C. Ressi's *Game Over* (2018) and G. Peles' *Score Craft* (2022). Additionally, research projects like "GAPPP: Gamified Audiovisual Performance and Performance Practice" led by M. Ciciliani [2] and the "Digital Score" project spearheaded by C. Vear [3] have contributed to the exploration and development of various gamified works.

The increased interest in gamified compositions stems from their potential to create emergent musical forms through performer-computer interactions. However, despite the growing prominence of these compositions, to the best of our knowledge, only one study has proposed a framework for identifying game elements used in musical compositions and analyzing how these elements contribute to playful performer-computer interactions [4]. This scarcity may be because other analytical frameworks focus primarily on identifying game elements in consumer products rather than musical compositions. This paper aims to address this gap by proposing the Expanded Motivational Affordances (EMA) model as an analytical framework that identifies game elements and examines how the elements support playful performer-computer interactions in gamified screen-score works.

In the following sections, we introduce two crucial analytical frameworks: the Mechanics-Dynamics-Aesthetics model and the Motivational Affordances model. Their comparison underscores the necessity of expanding the Motivational Affordances model, which serves as the cornerstone of our research. Moving forward, we present an overview of the EMA model in the third section. In the subsequent fourth section, we conduct a case study, employing the EMA model to analyze two gamified screen-score works: *Super Colliders* by T. Fukuda and *SQ2* by P. Turowski. This case study yields a comprehensive list of game elements, demonstrating the effectiveness of the EMA framework in this regard. Following this, we delve into discussions regarding implications and limitations in the fifth section, outline potential directions for future research, and conclude by summarizing the validity of the EMA framework for understanding and harnessing the potential of the gamified screen-score works.

2. RELATED WORKS

There have been several frameworks that aid in understanding a game in the field of game design [5, 6, 7]. When it comes to creating engaging interactions between players and computers, one of the most pertinent frameworks is the Motivational Affordances framework.

[1] Interactive computer music refers to "a music composition or improvisation where software interprets a live performance to affect music generated or modified by computers" [2]

[2] http://gappp.net

[3] https://digiscore.github.io

2.1 Motivational Affordances

The Motivational Affordances (MA) framework is founded on the MA theory, which postulates that enjoyable human-computer interactions depend on fulfilling users' motivational needs [8, 9]. The term "motivational affordances" refers to the properties of an object that determine whether and how it can support users' motivational needs. Based on the MA theory, Weiser et al. have established a framework that serves as an analytical tool to identify several game elements and how these elements fulfill various motivational needs within gamified products [10]. Their taxonomy of motivational needs encompasses the following six categories:

- Autonomy: The need for the freedom to initiate and regulate one's own behaviour, making choices independently.
- Competence: The need to excel in mastering challenges appropriate for personal growth and development.
- Relatedness: The desire for a sense of belonging and connection to others.
- Achievement: The desire to perform exceptionally well, measured against a standard of excellence, which can involve competing with oneself, others, or specific tasks.
- Affiliation and Intimacy: The need for other people's approval and the inclination toward secure and rewarding relationships, respectively.
- Leadership and Followership: The desire to gain authority and impact, control, and influence others and the desire to support or be subordinate to a leader, respectively.

Weisler et al.'s study introduces the concept of "Mechanics," which are defined as "possible means of interaction between a user and the system" [10]. Engaging interactions are achieved when mechanics are designed to fulfill the user's motivational needs. The six key mechanics are identified:

- Feedback: Providing users with visual and aural information about their actions, optimizing their actions, and increasing motivation.
- User Education: Offering advice to compensate for knowledge gaps and assisting users in achieving their goals.
- Challenges: Presenting difficult tasks or quests that fulfill the desire for competence.
- Rewards: Offering valuable items or incentives in exchange for user accomplishments, satisfying achievement and competence needs.
- Competition and Comparison: Creating situations involving challenges and rivals, addressing the desire for achievement and leadership, often seen in multiplayer settings.
- Cooperation: Facilitating collaborative actions with other players to achieve common goals, addressing the need for affiliation and leader-/followership.

This framework forms the foundation for comprehending how the fulfillment of motivational needs through mechanics can foster the playfulness of performer-computer interactions in gamified environments. Finally, their study defines the term "game elements" as specific tasks or objects that support mechanics, such as quests (for Challenges), points (for Rewards), and leaderboards (for Competition).

2.2 MDA

The Mechanics-Dynamics-Aesthetics (MDA) model is another prominent tool for the analysis of computer games and has been considered in this context because it is a seminal framework for "playcentric" approaches.[6, 7] In other words, focusing on the experience of the player first and foremost "encourages experience-driven (as opposed to feature-driven) design."[11]

The MDA framework offers a comprehensive breakdown of games into three fundamental components:

- Mechanics: The fundamental building blocks of the game encompassing its rules, basic player actions, algorithms, and data representation within the game engine. Mechanics essentially define the core operations and functioning of the game.
- Dynamics: The real-time behaviour of these mechanics when players interact with them. It encompasses how the mechanics respond to player input and how they interact with one another throughout gameplay.
- Aesthetics: The desirable emotional responses elicited in players as they engage with the game.

The MDA framework further provides a taxonomy of aesthetic types that encompass a wide range of player experiences such as the following;

- Sensation: This aesthetic type portrays the game as a source of sensory pleasure, where players delight in memorable audio-visual effects.
- Fantasy: In this category, games offer a realm of make-believe, immersing players in imaginary worlds.
- Narrative: Games take on the role of drama, enticing players with engaging stories that keep them returning for more.
- Challenge: This aesthetic kindles the desire to conquer, presenting games as obstacle courses that fuel the urge to master them and enhance replayability.
- Fellowship: Games act as social frameworks, forming communities where players actively participate. This aesthetic is particularly prevalent in multiplayer games.

- Discovery: Games invite exploration, presenting themselves as uncharted territories that spark players' curiosity to uncover the game world's secrets.

- Expression: This aesthetic encourages self-discovery and creativity, allowing players to express themselves, such as creating characters resembling their own avatars.

- Submission: In this category, games serve as pastimes, fostering a connection to the game as a whole, even within certain constraints.

The MDA framework, with its categorization of aesthetics and analysis of game components, provides insights into the multifaceted nature of player experiences in computer games. Understanding these components aids in not only game design but also the enhancement of player engagement and enjoyment. The MDA framework stands as a tool for dissecting and comprehending the relationships between mechanics, dynamics, and aesthetics within computer games, contributing to the creation of more immersive and captivating gaming experiences.

2.3 Comparision Between MA and MDA Frameworks

In our pursuit of a comprehensive understanding of gamified screen-score works, we have conducted a comparison between two influential frameworks: the MA framework and the MDA framework. This comparative analysis has unearthed connections between these frameworks, shedding light on how each aesthetic type proposed in the MDA framework can effectively address the motivational needs outlined in the MA framework, as illustrated in Table 1.

For instance, consider the aesthetic type of "Fantasy". This aesthetic type has the capacity to address the need for autonomy because the imaginative immersion in a make-believe world hinges on players' freedom to make decisions within that realm ("1" in Table 1) [12]. Similarly, the aesthetic type "Challenge" can serve as a means to fulfill the desire for competence. The imposition of difficult challenges upon players contributes to the enhancement of their skill levels, aligning with their intrinsic desire for personal growth and proficiency ("5" in Table 1).

Furthermore, during our comparative analysis, we observed an intersection between Zhang's MA theory [9] and the MDA framework. Specifically, the "affect and emotion" category, representing emotional states that occur as reactions to significant stimuli in one's environment, in Zhang's motivational taxonomy is particularly relevant to the "Sensation" aesthetic type in the MDA framework. This association stems from the fact that the sensation aesthetic type often revolves around creating sensory pleasure, which can shape player emotions. In light of this observation, we made a decision to augment the taxonomy of motivational needs initially proposed by Weiser et al [10] by incorporating the "affect and emotion" category.

In the subsequent section, we will introduce our Expanded Motivational Affordances model and elucidate how these refinements augment our understanding of the intricate relationships between aesthetics, mechanics, dynamics, and motivational needs within gamified screen-score works.

3. THE EMA MODEL

3.1 Our Model

The Expanded Motivational Affordances (EMA) model offers an intricate framework designed to comprehensively grasp the dynamics underpinning user engagement in gamified screen-score works. This model recognizes and delves into a diverse array of motivational needs, which users actively seek to fulfill while immersing themselves in these interactive musical compositions. These motivational needs encapsulate a wide spectrum of physiological, psychological, and social desires, collectively constituting the prime motivators that drive user engagement and participation.

In both the MA and MDA frameworks, prioritizing an understanding of the player's experience is pivotal for identifying mechanics that effectively achieve specific outcomes crucial for gamified composition designers. In other words, both frameworks underscore the importance of 'experience-driven design,' prompting a desire to integrate these two models.

3.2 Components of the EMA Model

The EMA model is structured around a taxonomy of motivational needs, each representing a distinct facet of the user's intrinsic drive for engagement. These motivational needs are as follows:

- Autonomy: This fundamental need pertains to the desire for autonomy, empowering users with the freedom to initiate and regulate their behaviour within the gamified screen-score experience. It encompasses the ability to make independent choices and decisions, thereby fostering a sense of agency. Affording autonomy contributes to imaginative immersion (fantasy and narrative), personal expression and emotional maintenance (submission).

- Competence: Users are innately driven by the pursuit of excellence and the desire to master challenges that are optimally suited for their personal growth and development. The competence need fuels the aspiration for continuous improvement. Competence is also related to the process of discovery—i.e., understanding the breadth and scope of the system.

- Relatedness: At its core, the human experience thrives on social connections. The relatedness need encompasses the yearning for social engagement, recognition, and a deep sense of belonging within the community of participants. Integrating fantasy and narrative can help players to relate to their environment, and submission can also be supported by fellowship.

- Achievement: Individuals aspire to showcase their competence and accomplishments to others. The

		Motivational needs						
		Autonomy	Competence	Relatedness	Achievement	Affiliation and Intimacy	Leadership and Followership	Affect and Emotion
Aesthetic types	Sensation							17
	Fantasy	1		7		13		18
	Narrative	2		8				
	Challenge		5		11		15	
	Fellowship			9		14	16	
	Discovery		6		12			
	Expression	3						
	Submission	4		10				

Table 1. The table shows which motivational needs each aesthetic type may satisfy.

achievement need underscores the motivation to excel and be recognized for one's capabilities and achievements. Rising to a challenge and discovering more of the available possibility space can support a sense of achievement.

- Affiliation and Intimacy: Human connections are nourished by approval, affection, and the cultivation of secure and rewarding relationships. This need encompasses the desire for social approval, as well as the inclination towards forming intimate and meaningful bonds, which can contribute to a sense of immersion.
- Leadership and Followership: Users may seek to exert influence and control over their surroundings, aspiring to shape the physical or social environment in alignment with their vision or plan. Conversely, there is also a desire to support or be subordinate to a leader, reflecting the leadership and followership needs. Effective team management can satisfy the desire to overcome challenge.
- Affect and Emotion: Recognizing the influence of emotions on the user experience, this category emphasizes induced emotional states that arise in response to significant stimuli. It encompasses the emotional responses that play a pivotal role in shaping the overall motivational experience, enriching it with affective dimensions.

The EMA model extends and enriches the foundational concepts of the MA framework. By incorporating the "affect and emotion" category into the taxonomy of motivational needs, the EMA model provides a more comprehensive and nuanced perspective on the factors influencing player-computer engagement in gamified screen-score compositions. This expansion equips researchers, designers, and practitioners with a refined tool for identifying and understanding the game elements that contribute to the multifaceted motivational experiences within this genre of interactive computer music.

The subsequent section of this paper is dedicated to in-depth case studies of existing gamified screen-score works. These case studies employ the EMA model as a lens to comprehensively identify various game elements within two existing gamified screen-score compositions: *Super Colliders* and *SQ2*.

4. CASE STUDIES

4.1 *Super Colliders*

4.1.1 Goals and Rules

Super Colliders (2018) is a composition created for three pitched instruments and an interactive computer system. The primary objective was to craft the piece in such a way that the player's desire to win the game results in an enriching musical performance [4]. Therefore, it can be assumed that the piece aims to satisfy the motivational needs of competence and achievement. The interactive computer system poses challenges as musical symbols, which the players must interact with to perform.

The game consists of four rounds in which players compete against each other. At the end of the game, one player is declared the grand winner based on their success in winning the most rounds. To win each round, players must accumulate 1,080 life points faster than their opponents. These life points are earned by colliding their avatars with continuously moving blobs on the screen-score. Each collision results in the acquisition of one life point. However, if all players miss a collision, all three of them lose a point collectively. Thus, the game has two possible outcomes: either one musician earns the highest number of points and wins the round, or a scenario called "all dead" occurs, signifying that all musicians lose and the computer emerges victorious.

During the performance, the players control their avatars by executing ascending or descending glissandi at varying volumes. The vertical position of the performers' avatars corresponds to the pitch change, while the horizontal position and avatar size are mapped to the loudness of their performance [13].

4.2 Identification of Game Elements

In gamified screen-score compositions, the mechanics that drive performer-system interaction find concrete expression through design components referred to as "elements." These elements, specific to each gamified piece, support and manifest the underlying mechanics, translating abstract concepts into tangible user experiences. While Weiser et al.[10] originally identified seven universal and context-free element categories, this subsection delves into the context-specific instantiation of mechanics as concrete elements with the purpose of discovering new elements specific to gamified screen-score works.

4.2.1 Feedback

Feedback mechanisms in this context include:

- Obedient avatars.
- Responsive collision sounds.

Obedient avatars and collision sounds serve as immediate and perceptible feedback to players regarding their performance within the screen-score. This feature inherently caters to the need for self-determination during the performance, thereby affording the motivational need for autonomy. Moreover, the predictability of avatars' reactions to sound effects enhances the performers' sense of control, contributing to a sense of competence, especially when fully immersed in gameplay.

4.2.2 User Education

Elements facilitating user education encompass:

- Written instructions.
- Rehearsals.
- Rounds.

These elements bridge the knowledge gap for performers, elucidating the requirements of the performance. They support the competence need by equipping performers with the necessary knowledge and skills to engage effectively. Moreover, if competence leads to success and achievement, user education plays a crucial role in fulfilling the achievement need. Written instructions provide a conceptual framework for the piece and aid in performance preparation. Rehearsals offer valuable opportunities for performers to familiarize themselves with the interactive system, fostering skill development and strategic thinking. The iterative nature of rounds further serves as a learning platform, allowing performers to refine their skills and interpretations as the piece unfolds.

4.2.3 Challenges

Challenges are embodied through:

- Moving blobs.

The task of intercepting moving blobs with avatars offers a challenging experience that caters to the competence need. Success in this task demands vigilance, visual acuity, and agility, satisfying performers' desire for mastery. Striking a balance between challenge and frustration is crucial, with blob behaviours serving as the key factor in controlling difficulty levels. These challenges are pivotal in maintaining engagement throughout the piece.

4.2.4 Rewards

Reward elements consist of:

- Life points.

Life points serve as a mechanic to reward players for successfully intercepting blobs, aligning with the competence need. Additionally, they fulfill the achievement need by visibly showcasing the player's success. Life points introduce a competitive aspect, where players' health is compared, potentially leading to 'deaths' within the game. This concept fosters both competition and cooperation dynamics among players, making them view each other as rivals and strategic collaborators to avoid 'dying.' Thus, life points satisfy the need for survival, intricately linked to leadership and followership.

Figure 1. The screen-score showing avatars, moving blobs and life point indicators in *Super Colliders*

4.2.5 Competition

Competition is epitomized by:

- Leaderboard.

The leaderboard announces round winners and, ultimately, the grand winner at the conclusion of the piece. This element resonates with the relatedness and achievement needs, providing a platform for recognition and social engagement among performers.

4.2.6 Cooperation

Cooperation manifests through:

- 'All dead' scenario.

While primarily a competitive setting where only one player can win, the 'all dead' scenario presents an incentive for players to cooperate and prevent missing too many blobs, thus depleting their life points. This element addresses the needs for leadership and followership within the ensemble, emphasizing collaboration among players even in a competitive environment.

4.3 *SQ2*

4.3.1 *Goals and Rules*

SQ2 (2019) may be played by any quartet of performers that are capable of producing sustained pitch with at least a one-octave range and varying degrees of loudness.

In *SQ2*, players take control of four luminescent orbs. When a player makes a sound, their corresponding orb creates circular pulses of energy, which collide with nearby objects and agents. By sustaining pitches, energy pulses can be tuned to resonate with particular targets (see Fig. 2). Players may create energy pulses as long as they have stamina remaining; if a player's stamina is fully depleted, they must wait until it fully recharges before creating another pulse.

Players work together to shape the sounding score and facilitate growth within the virtual environment—e.g. resonating with colored gems (i.e. matching pitch class) to produce tree-like structures. While some elements of the score—e.g. overall form—are prescribed, players are free to decide how lower-level patterns—e.g. rhythm, melody, rate of progression—will develop over time. The game ends when the performers collectively maintain a long period of silence.

SQ2 primarily focuses on collaborative gameplay and therefore aims to satisfy the motivation needs for leadership and followership as well as affiliation and intimacy, emphasizing fellowship among players [14].

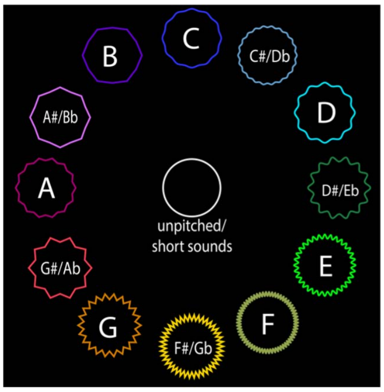


Figure 2. Mapping of shape/color to pitch class in SQ2

4.4 Identification of Game Elements

4.4.1 *Feedback*

Feedback mechanisms in this context include:

- Obedient avatars.
- Electronic sounds as a pitch and rhythmic reference.
- UI meters.

Feedback in *SQ2* is designed to be both 'discernable' and 'integrated', as key aspects of 'meaningful play'[5], which relates to autonomy and competence. Feedback is discernable through immediate visual cues on successful control input in the form of an audio signal. These visual cues are redundant for the sake of greater visibility and accessbility; input is registered in the UI meters at the bottom of screen as well as on the player's avatar, and an extensible line connects the two in later stages to avoid confusion. Furthermore, meters display information about pitch class, available stamina (graphical and numerical), and the time threshold for energy circle creation. (See Fig 3.) At certain points in the progression of the score, the software produces electronic pitches at regular intervals that provide a framework (pitch and rhythm) for improvisation. Feedback is 'integrated' in that "an action a player takes not only has immediate significance in the game, but also affects the play experience at a later point in the game" [5]. This is achieved through the transformation of the environment over the duration of the piece.

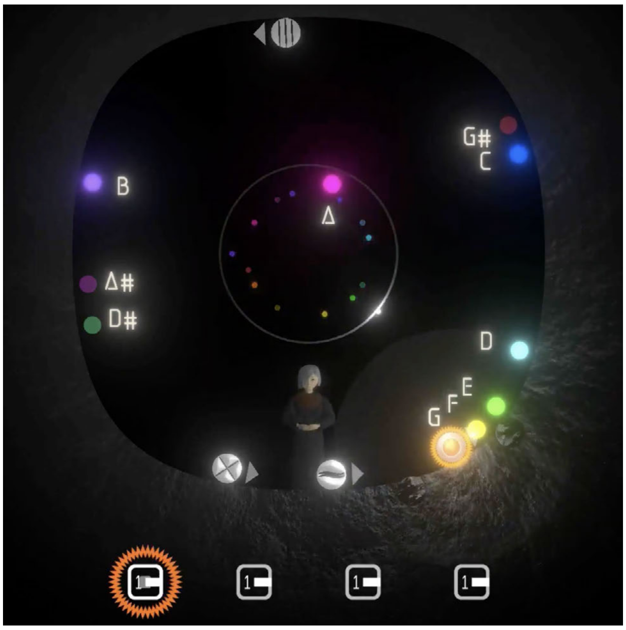


Figure 3. SQ2 user interface

4.4.2 *User Education*

Elements facilitating user education encompass:

- Didactic design.
- Written instructions.
- Instruction video.
- Rehearsal.

The piece requires the completion of basic tasks in the beginning that introduce the players to the mechanics of the game and gradually increase in complexity. Initially, any sound event (a sound that lasts for more than 100ms and terminates in silence) creates an energy pulse that moves the player's avatar forward. (The forward vector is indicated by a small arrow near the avatar.) Energy pulses also release gems from the terrain, and players can resonate with these gems by playing particular pitch classes. Resonating with a gem increments the collective point total, and achieving the target score allow players to progress to the next stage. These basic 'control sound' [15] mechanics are the foundation for more complex mechanics that are revealed later, such as tree growth on the surface of the planet. Defining a 'core mechanism' to achieve a more 'elegant' design [16] serves a didactic purpose and therefore contributes to the motivational need of competence. Education towards a greater sense of competence is further supported by written performance instructions and a tutorial video [17], which should inform the rehearsal process. Given the improvisatory and collaborative nature of the piece, rehearsals are crucial for a proper performance of this piece.

4.4.3 Challenges

Challenges are embodied through:

- Time limit (optional).
- Agents.

SQ2 was intended to be an open exploratory experience with a relatively low level of challenge (after the initial learning phase). For example, players can take as long as they want to meet the point-based goals in each stage. (Points and the related stamina system are further discussed in the 'Rewards' section below.) However, an optional time-limit allows for the length of performances to be more strictly regulated and this can increases the level of challenge by giving players an extra task—i.e. to satisfy the point goals before time runs out. Additionally 'agents' or NPCs (non-player characters) roam the terrain and attract players who are making non-pitched sound. Agents can either restrict the player (if they wish to avoid being attracted) or facilitate movement (by changing direction or by 'riding' an agent to a desired location on the map). In either case, players benefit from being aware of agent positions, which can be difficult when focusing on the core tasks of movement and gem resonance. Both elements add challenge to the play experience and thus address the motivational needs of competence and achievement.

4.4.4 Rewards

Reward elements consist of:

- Stamina extension.
- Points.
- World development.

Successful achievement of the primary game goal (i.e. to resonate with colored gems) is rewarded in three ways. Resonating with gems increases the overall point tally, which provides players with a clear and quantifiable sense of collective progression through the piece. Additionally, gem resonance incrementally increases player stamina, which governs how many energy pulses a player can create. In other words, having more stamina allows a player to act (i.e. move and resonate) more continuously rather than having to wait in silence for stamina to replenish. This reward satisfies the motivational need for autonomy. Lastly, progression yields a transition from the 'inner earth' environment to the 'planet surface' environment and subsequently populates the surface with tree-like structures, which modify the traversable terrain and establish the harmonic palette and rhythmic period of the resulting improvisation. All three of these reward types support the motivational need of achievement.

4.4.5 Competition

This piece is designed to avoid competitive gameplay. Players can elect to stifle each other if they wish, but such behaviour would delay/prevent the achievement of the goal to reach the indicated point score, which is shared between all players (see 'Cooperation' below). Players could potentially compete to see who can achieve the highest stamina level, but the mechanics of the game do not explicitly reward such behaviour.

4.4.6 Cooperation

Cooperation manifests through:

- Shared score.
- Player-to-player stamina replenishment.

Resonating with colored gems is rewarded by gaining a point in a collective point system, which regulates the overall progression of the piece and encourages collaborative behaviour to satisfy the motivational needs of relatedness, affiliation and intimacy, and leadership and followership. Additionally, a player is able to help another by creating an energy circle near them, which replenishes their stamina.

5. COMPARISON BETWEEN *SUPER COLLIDERS* AND *SQ2*

The most notable difference between *Super Colliders* and *SQ2* is that the former focuses on a competitive gameplay experience while the latter focuses on collaborative gameplay experience. While both satisfy most of the seven motivational needs to some extent, Competence and Achievement are more primary for *Super Colliders* while Relatedness and Leadership and Followership are more primary for *SQ2*. These motivational differences necessarily result in mechanical differences, such as the different implementations of common game elements like points, agents, and time limits.

6. DISCUSSIONS

The introduction of the Expanded Motivational Affordances (EMA) framework presents an exciting development in the analysis and creation of gamified screen-score compositions within the realm of interactive computer music. In this discussion, we introduce the implications, limitations, and future directions of the EMA framework.

6.1 Implications

The EMA framework offers profound implications for both research and composition in the field of gamified screen-score works.

Firstly, as demonstrated through our case studies of *Super Colliders* and *SQ2*, the EMA framework provides a powerful lens for understanding existing compositions. By dissecting these works into their constituent game elements and linking them to specific motivational needs, we gain insights into the intricate interplay between performers and computer systems. This understanding not only enriches our appreciation of these compositions but also lays the foundation for more informed critique and analysis.

Secondly, the EMA framework emerges as a potential model for the creation of new gamified screen-score works. Composers can leverage this framework to strategically design and integrate game elements that cater to desired motivational needs. This approach opens up exciting possibilities for the deliberate fusion of musical and gaming elements, fostering innovative and engaging interactive musical experiences.

6.2 Limitations

While the EMA framework holds significant potential, it is essential to acknowledge its limitations.

One notable limitation is the potential contextual dependency of motivational needs. The degree to which specific game elements fulfill these needs may vary across different cultural and contextual settings. Therefore, the framework's universal applicability should be approached with caution. Future research should explore cross-cultural studies to better understand these variations.

Another limitation concerns the need for empirical validation. While the framework is theoretically grounded, empirical studies are necessary to confirm its effectiveness in analyzing and designing gamified screen-score compositions. Such studies could involve performer feedback, audience response analysis, and comparative assessments of compositions that utilize the EMA framework versus those that do not.

We have intentionally limited the scope of this paper to focus primarily on the application of game theories to gamified works and higher level decisions about how creators design mechanics. The relationship between musical criteria and the EMA framework is beyond the scope of this paper and may be the subject of future inquiry.

6.3 Future Work

Future work in this field should prioritize several key directions:

Empirical Validation: Rigorous empirical studies should be conducted to validate the EMA framework. This includes gathering feedback from performers and audiences engaged with gamified screen-score compositions analyzed using the EMA model. This feedback can shed light on the framework's effectiveness and areas for improvement.

Cross-Cultural Studies: To account for potential contextual variations, cross-cultural studies are essential. Investigating how the EMA framework operates in diverse cultural contexts can provide valuable insights into the framework's adaptability and cultural sensitivity.

Interdisciplinary Collaborations: Collaboration between composers, musicians, game designers, psychologists, and human-computer interaction experts can further enrich the EMA framework. Interdisciplinary approaches can foster innovative research and composition methodologies that push the boundaries of gamified screen-score works.

7. CONCLUSIONS

In this paper, we have introduced the Expanded Motivational Affordances (EMA) framework as an analytical tool for the study of gamified screen-score compositions in the field of interactive computer music. The EMA framework expands upon the Motivational Affordances model by incorporating the category of "Affect and Emotion" within a taxonomy of motivational needs. This expansion provides a comprehensive lens through which to identify and understand the game elements that underpin performer-computer interactions in these compositions.

Through case studies of gamified screen-score compositions, specifically *Super Colliders* by T. Fukuda and *SQ2* by P. Turowski, we have exemplified how the EMA framework can be applied to identify and categorize game elements that cater to diverse motivational needs. These elements range from obedient avatars, responsive collision sounds, rounds, and leaderboards to the 'all dead' scenario, showcasing the richness of the game elements interfacing between performer and computer in the interactive system design of these works.

The implications of the EMA framework are twofold. Firstly, it offers a valuable lens for understanding existing gamified compositions, shedding light on the intricate motivations that drive performer engagement. Secondly, it serves as a potent model for the creation of new gamified screen-score works, guiding composers in the strategic incorporation of game elements to enhance the performer's engagement with the gamified computer system.

However, we acknowledge certain limitations of the EMA framework, such as its potential contextual dependencies and the need for further empirical validation through an analysis of more works as samples. As such, future work in this field should focus on empirical studies to refine and expand the framework's applicability.

In conclusion, we believe that the Expanded Motivational Affordances framework represents a significant step forward in understanding and harnessing the potential of gamified screen-score compositions. It offers an analytical lens that emphasizes the fulfillment of motivational needs as a driving force behind the playfulness and engagement that define this evolving genre of interactive computer music. With continued research, refinement, and interdisciplinary collaboration, the EMA framework holds the promise of enriching both the study and creation of gamified screen-score works, paving the way for innovative and captivating compositions within the musical community in the digital age.

Acknowledgments

We would like to thank the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), the Interdisciplinary Centre for Composition and Technology (ICCaT), Concordia University and the University of Liverpool for their support.

8. REFERENCES

[1] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining "gamification"," in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, ser. MindTrek '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 9–15. [Online]. Available: https://doi.org/10.1145/2181037.2181040

[2] T. Winkler, *Composing interactive music: techniques and ideas using Max*. MIT press, 2001.

[3] M. Ciciliani and B. Lüneburg, "Gamified audiovisual works–composition, perception, performance," *Emille, the Journal of the Korean Electro-Acoustic Music Society*, vol. 15, pp. 21–26, 2018.

[4] T. Fukuda, "Super colliders: A gamified screen-score," in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR'20/21*, R. Gottfried, G. Hajdu, J. Sello, A. Anatrini, and J. MacCallum, Eds. Hamburg, Germany: Hamburg University for Music and Theater, 2020, pp. 158–168.

[5] K. Salen Tekinbas and E. Zimmerman, *Rules of Play: Game Design Fundamentals*. Cambridge: MIT Press, 2003.

[6] T. Fullerton, *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*, 4th ed. United States: A. K. Peters, 2018.

[7] J. Schell, *The Art of Game Design: A Book of Lenses*. United States: A. K. Peters, 2019.

[8] P. Zhang, "Toward a positive design theory: Principles for designing motivating information and communication technology," in *Designing Information and Organizations with a Positive Lens*, ser. Advances in Appreciative Inquiry, M. Avital, R. J. Boland, and D. L. Cooperrider, Eds. Emerald Group Publishing Limited, dec 2007, vol. 2, pp. 45–74. [Online]. Available: https://doi.org/10.1016/S1475-9152(07)00204-9

[9] ——, "Technical opinion: Motivational affordances: Reasons for ict design and use," *Commun. ACM*, vol. 51, no. 11, p. 145–147, nov 2008. [Online]. Available: https://doi.org/10.1145/1400214.1400244

[10] P. Weiser, D. Bucher, F. Cellina, and V. D. Luca, "A taxonomy of motivational affordances for meaningful gamified and persuasive technologies," in *Proceedings of EnviroInfo and ICT for Sustainability 2015*. Atlantis Press, 2015/09, pp. 271–280. [Online]. Available: https://doi.org/10.2991/ict4s-env-15.2015.31

[11] R. Hunicke, M. Leblanc, and R. Zubek, "MDA: A formal approach to game design and game research," *AAAI Workshop - Technical Report*, vol. 1, 01 2004. [Online]. Available: https://api.semanticscholar.org/CorpusID:6056852

[12] L. Ermi and F. Mäyrä, "Fundamental components of the gameplay experience: Analysing immersion." 01 2005.

[13] T. Fukuda, "Super Colliders performance," April 2018. [Online]. Available: https://vimeo.com/382978696

[14] P. Turowski, "SQ2 performance," Nov. 2019. [Online]. Available: https://youtu.be/amlC2U5zTP8?si=zcgz1kv5pv72PXNA

[15] ——, "Digital game as musical notation," Ph.D. dissertation, University of Virginia, 2016. [Online]. Available: https://doi.org/10.18130/V3HS18

[16] K. Burgun, *Clockwork Game Design*. CRC Press, 2015.

[17] P. Turowski, "SQ2 tutorial," Sep. 2022. [Online]. Available: https://youtu.be/BC3caaiGXjk?si=YSFNGT2xoxD0hKAH

TABstaff+: A HYBRID MUSIC NOTATION SYSTEM FOR GRID-BASED TANGIBLE USER INTERFACES (TUIs) AND GRAPHICAL USER INTERFACES (GUIs)

Lawrence Wilde
University of Siegen
Faculty II: Education · Architecture · Arts
lawrence.wilde@uni-siegen.de

Charles White
Max Weber Center for Advanced Cultural and Social Studies at University of Erfurt and University of Graz
charles.white@uni-erfurt.de

ABSTRACT

TABstaff+ is a hybrid music notation developed for grid-based user interfaces. The system builds on notational elements and conventions of tablature, standard five-line staff notation, and chord diagrams. TABstaff+ strives to facilitate teaching and learning, composition and production, and performance using grid-based tangible user interfaces (TUIs) and graphical user interfaces (GUIs). For usability testing, the study involved seven participants, music production and composition students (ages 13 to 19) with prior musical experience. The paper considers the Ableton Push instrument to illustrate the application and adaptability of the TAB+, Staff+, and Charts+ notation systems. These notation systems aim to further the development of postdigital practices by leveraging Human-Computer Interaction (HCI) and pre-digital practices of reading, playing, and teaching music using instruments and notation. TABstaff+ aims to be a transferable music notation system that allows educators and practicing musicians to utilize the pedagogical and creative capabilities of musical grid interfaces.

1. INTRODUCTION

Musical Instrument Digital Interface (MIDI) grid instruments and touchscreen graphical user interfaces (GUIs) have changed how music is composed, performed, taught, and learned [1] – [5]. Considering these developments, comprehensive and dedicated notation systems are needed for musical grid interfaces. As Giles [6] points out, 'due to the grid controller's primary association with more popular styles of music, a scoring paradigm for it has not been explored.' Many of the compositions and performances utilizing these types of interfaces, to a large extent, employ un-notated and improvisatory approaches. While many of these works are recorded via audio, MIDI, video, and other means, they tend to be fixed or semi-fixed media. A notation system for musical grid interfaces can enable the education, creation, and live performance of digitally produced works, leading to new educational paradigms, artistic possibilities, and forms of dissemination [7]. As musical instruments with grid interfaces offer 'an utterly unique method of controlling a computer system' [6], a notation system that accounts for grid playing surface topographies can significantly contribute to Human-Computer Interaction (HCI) in digital music practices, as the grid controller is becoming an industry standard and an "object of research in the HCI community" [8].

While digitization of music has enabled new creative and collaborative opportunities, there remains a gap in HCI in terms of reading, learning, teaching, and performing digitally created music [9, 10, 11]. The use of musical grid interfaces, like the Ableton Push, Novation Launchpad, Native Instruments Maschine, Akai MIDI grid controllers, and software applications featuring touchscreen grid GUIs such as Ableton Note, Groovepad, Drum Pad Machine, and others, have been a step forward in human-machine music making, enabling immediate and tactile engagement with digital material. These developments where already underway in the 1960's with "push-button matrices" that "were used in electronic devices (cf. phones) and research and development (US3676607A) created new technologies which influenced the design of later grid interfaces" [12]. In the late 1980's MPC controllers were introduced and quickly adopted by the music industry [13]. Advancements in touch-sensitive technologies have allowed for more expressive control over sound manipulation; parameters like velocity, MIDI Polyphonic Expression (MPE), finger position tracking, and others have empowered musicians to achieve unprecedented nuance and creative expression. These advancements invite new ways for fostering creative reusability and reproducibility (e.g., live performance) of digital musical material [14] – [16]. The challenge lies in translating digital musical information created through Digital Audio Workstations (DAWs) and other technologies into live performance contexts utilizing grid interfaces. Finding ways to bridge the gap between digital and physical environments remains essential for furthering music creation, performance, and education in the postdigital age.

A real-world example of this challenge in education can be a music teacher designing a class activity in which students learn fundamental elements of music by triggering of loops, controlling parameters of sound, and playing melodies and chords on grid TUIs or GUIs. Effectively communicating musical instructions poses a challenge due to the absence of a recognized notation system that graphically represents playing actions for grid instruments. One option is for the teacher to create their own notation system or set of instructions. While this may work for an isolated case, such a system may be limited in terms of reproducibility in different contexts, with other educators and students needing to 'decipher' the 'new symbolic language' [17, 18]. Furthermore, developing such a notation system is highly demanding for music educators, taking time away from teaching-related activities. While designing a hands-on classroom activity in the DAW may be relatively straightforward, the difficulty lies in finding a way to translate digital material into a human-readable format optimized for learning and playing on grid instruments [19].

Music educators and musicians who work with traditional and non-electronic mediums use various established notation systems. These include the standard five-line staff, tablature, chord diagrams, percussion-specific notation systems, and world music notation systems. These systems enable the clear communication of intricate musical concepts and effectively support activities such as teaching, learning, composing, and performing. Nevertheless, many systems, like tablature and chord diagrams, are often instrument-specific. Even general-purpose systems, like the five-line staff notation, are optimized for traditional instruments such as piano, violin, flute, and others [20]. Although these systems work well in their respective contexts, they are not as effective in representing and communicating musical material created digitally and intended for performance on musical grid interfaces. Graphical representations of musical material in DAWs exist in the form of the piano roll, timeline, MIDI data, and other visualizations. These are helpful for activities such as MIDI programming, sequencing, and automation in composition and production. However, they are not optimal for grid interface-based learning, teaching, and performance. Traditional DAW visualizations like piano rolls or MIDI data are often static and complex, which can be challenging for beginners to understand. Musical content created using MIDI programming in a DAW can be saved as a MIDI file, converted into MusicXML format, and imported into music notation software like Dorico, Finale or MuseScore, allowing for the digital music data to be represented in standard five-line staff notation or tablature for human readability in performance practices. However, traditional notation systems are not optimized for readability and playability on grid TUI and GUI instruments. These challenges can be overcome through the development of new notation systems designed specifically for musical grid interfaces offering intuitive, visual and tactile ways of learning music.

The TABstaff+ notation system introduced in this paper recognizes that 'notation of electronic music' is part of the 'evolutionary lineage' of notation systems and builds on the standard five-line staff, tablature (TAB), and chord chart notation conventions [21]. These established graphical and symbolic representations of music are utilized in developing TABstaff+ due to their widespread familiarity, usability, and transferability across multiple musical styles, genres, and practices. TABstaff+ encompasses three notation subsystems: TAB+, Staff+, and Charts+. These new notation paradigms combine elements from the traditional notation systems, expanding each to account for the unique layout of grid interfaces.

2. EDUCATIONAL IMPLICATIONS

National and international music curriculum guidelines underscore the importance of equipping students with the ability to 'use a system, e.g., staff notation or TAB, to learn and perform music appropriate to the instrument and musical style' [22]. Additionally, new teaching and learning frameworks emphasize digital aptitude and literacy in teacher training and at all levels of education [23]. As grid TUIs and touchscreen GUIs are being increasingly integrated into the music classroom [24] – [27], a notation system developed specifically for these types of interfaces can serve as a means of furthering the teaching of music and fostering digital skills. Research acknowledges 'that problems in music reading skills hold back countless students and may be a major cause for them to drop out of music lessons,' a systematic and user-friendly notation for grid-based instruments, already commonly used in many educational settings, has the potential to contribute to the development of musical and digital competencies [28], [29]. From the perspective of the music educator, it has been noted that 'many music teachers tend to abandon music reading instructions in the initial phase of their program or at least try to minimize the emphasis on music literacy' [28], [29]. However, understanding that 'fragmented musical knowledge results if reading is not taught hand-in-hand with playing the instrument' underscores the importance of musical instruments and notation for holistic music education [30] – [32]. As grid user interfaces (UIs) are designed to 'reduce learning' in terms of cognitive dissonance by 'eliminating confusion,' they can streamline the learning process [8], [12]. In this context, 'reduce learning' is viewed positively, as in minimizing the amount of new information or skills that a user must acquire to effectively use a system. For grid UIs, this might mean that the interface is designed in such a way that a user can intuitively navigate and use it without needing extensive instruction or trial and error. Grid interfaces are typically designed for clarity and ease of use: each element's purpose is clear, and there are no ambiguous controls or commands. They are created to match users' expectations and preconceived notions of functionality, which minimizes the likelihood of cognitive dissonance, where a user's experience clashes with their expectations. A comprehensive and user-friendly notation system for grid instruments holds the potential to support the development of musical aptitude and digital proficiencies.

3. METHODOLOGY

The development of TABstaff+ hybrid music notation system involved a multi-stage process, including (1) literature review, (2) Design-Based Research (DBR), and (3) usability testing with seven music students (ages 13–19). The study began by identifying where additional research was needed on musical notation for grid-based TUIs and GUIs. The notation system developed in this study drew on prior research (e.g., [6], [17]) on notation for instruments like Push and comparable grid-based musical interfaces. The literature review shaped the development of the conceptual framework for the TABstaff+ notation system, encompassing (1) adaptability across musical genres, (2) ease of use for educators and students, and (3) compatibility with various grid-based instruments and interfaces. The Design-Based Research (DBR) stage included (1) research of grid-based instruments, (2) analyses of musical notation systems, (3) initial design of TABstaff+, (4) prototyping and iteration, and (5) final TAB+, Staff+, and Charts+ notation systems. The three subsystems of TABstaff+: TAB+, Staff+, and Charts+, underwent multiple iterations for refining of graphical representations, symbols, and usability. To ensure compatibility and transferability of the notation, the prototypes were tested on various grid-based instruments and music applications featuring touchscreen grid GUIs. The usability testing involved seven private music composition and production students, ages 13 to 19, with varying musical abilities. Participant feedback and observations were collected during the testing phase. Students were asked to perform specific musical tasks (playing scales, playing chords, triggering loops, etc.) using TABstaff+ and provide feedback on their experiences. The observations and feedback were used to refine the TABstaff+ notation system and contributed to the development of the final TAB+, Staff+, and Charts+ notation subsystems.

4. NOTATING FOR THE PUSH

"It falls to the educator to [...] cater for as wide a range as possible. And I think there is a portion of students who, upon learning to play the Ableton Push, would greatly benefit from a formal notation system to scaffold and guide their practice. This is not to say that Ableton educators are not aware of this and acting on it daily, but that a notation system would surely add to the variety of methods available to learn the instrument in a guided and supported way" [17].

The challenge of creating a notation system for the Push and similar musical grid interfaces lies in the numerous layout configurations afforded by MIDI mappings. Initiating the Push in Live, the user is presented with options for setting the topography of the instrument at the *mega*, *macro*, and *micro* level layouts. The layouts exhibit a nested hierarchical structure. The mega layout includes two modes: 'Note' and 'Session.' One can switch between these using buttons with corresponding names and symbols on the Push. Within the 'Note' mode, seven *macro* layouts are available:

- Melodic: 64 Notes
- Melodic: Sequencer
- Melodic: Sequencer + 32 Notes
- Drums: 64 Pads
- Drums: Loop Selector
- Drums: 16 Velocities
- 'Session' Layout

The *macro* layouts contain the following *micro* layouts:

- 4ths 'In Key' Vertical
- 3rd 'In Key' Vertical
- Sequent 'In Key' Vertical
- 4ths 'Chromatic/ Vertical
- 3rd 'Chromatic' Vertical
- Sequent 'Chromatic' Vertical
- 4ths 'In Key' Horizontal
- 3rd 'In Key' Horizontal
- Sequent 'In Key' Horizontal
- 4ths 'Chromatic' Horizontal
- 3rd 'Chromatic' Horizontal
- Sequent 'Chromatic' Horizontal [17, 33, 34].

The micro layouts can be set using the mode, layout, and direction parameters, adjusted using the encoders (knobs) and buttons above and below the display screen. The mode parameter contains two layout options: 'In Key,' in which only the notes that belong to the selected key are lit and accessible on the pads, and 'Chromatic,' in which all chromatic notes are available, with the notes within the scale lit and those outside unlit [33]. The layout parameter contains three options: 4ths, 3rds, and Sequent. In the 4ths view, each pad is positioned a fourth higher from the pad row below; in the 3rds view, each pad is positioned a third higher, and in the sequent view, each pad is positioned an octave above. The direction parameters contain two further layout options: Vertical, in which scales are played from left to right, and Horizontal, in which scales are played from bottom to top.

While the above pertains to the Push, similar complex MIDI mappings are possible on other musical grid interfaces. The variety of layout choices showcases the versatility of grid controllers as dynamic musical instruments. However, it presents a challenge in terms of designing a notation system capable of accommodating diverse playing surface topographies. To facilitate the development of such a notation system, TABstaff+ introduces *nano* layouts that divide grids into smaller units.

5. TAB+: A TABLATURE SYSTEM FOR GRID INTERFACES

5.1. Single-Line Notation in TAB+

The 64-pad layout of the Push allows for the division of the pads into four 16-pad sets, *nano* layouts, each with a 4×4 matrix. The TAB+ notation introduces four new clefs, **A**, **B**, **α**, and **β** (pronounced as A-clef, B-clef, alpha-clef, and beta-clef), for specifying the nano layouts. The A and B clefs correspond to the two 16-pad nano layouts in the lower half of the instrument, while the α and β clefs correspond to the two nano layouts in the upper part (Fig. 1).

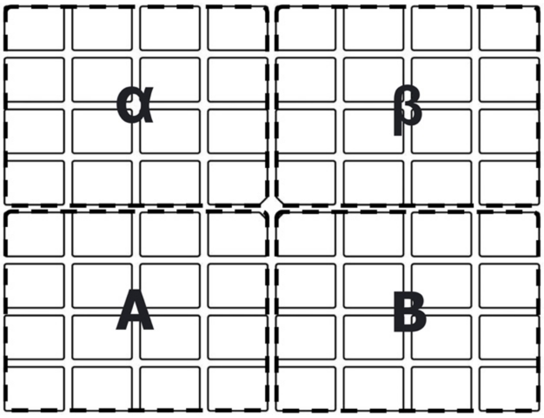


Figure 1. The main 64-pad grid playing surface of the Push divided into four 16-pad groups and labeled with TAB+ clefs **A**, **B**, **α**, and **β**.

The 4×4 matrices facilitate the development of the TAB+ four-line tablature system. This notation resembles traditional tablature systems; however, where lines usually indicate strings and numbers indicate frets, in TAB+, lines represent pad rows, and numbers specify pads within rows (Fig. 2). The adoption of elements from traditional tablature systems in the development of the TAB+ allows for the retention of key notational features such as the representation of time, note duration, rhythm, chords, contour, and other musical elements.

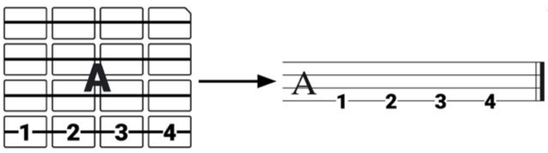


Figure 2. Four-line tablature with lines representing pad rows and numbers specifying pads in each row.

The TAB+ system allows musical structures such as melodies and harmonies to be read and played with ease. For example, in the default mode on the Push, an ascending one-octave C major scale (C2–C3) can be notated using the **A** clef. As the clef indicates, the scale is played in the bottom left nano layout (Fig. 3).

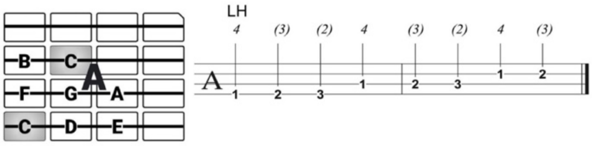


Figure 3. An ascending one-octave C major scale (C2–C3) notated in **A** clef for the left hand.

The TAB+ system incorporates note stems and beams for clarity of rhythm; however, it is also possible to use the system without stems and beams. The non-italic numbers on the tablature lines indicate the pads to be played, and the italic numbers above provide fingerings. To play a two-octave scale, the hand must traverse two nano layouts; there are two ways to notate this using the TAB+ system. For example, an ascending two-octave C major scale starting on C2 can be written using A and α clefs (Fig. 4 a) or only the A clef with ledger lines representing the transition into the pad rows of α (Fig. 4 b). Like the function of traditional clefs, the grid-specific alpha and beta clefs (i.e., A, B, α, and β) provide a reference point for interpreting the notes and their ranges. The TAB+ system further incorporates ledger lines as extensions of the tablature, resembling the function of ledger lines in the standard five-line staff notation system. This allows for the notation of musical material that extends beyond a single octave or calls for two or more nano layouts. Traditional tablature systems generally do not use clefs or ledger lines; this is a novel feature of the TAB+ notation system.

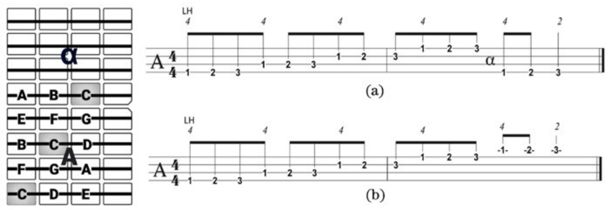


Figure 4. (a) An ascending two-octave C major scale (C2–C4) notated in **A** and **α** clefs for the left hand; (b) the same scale notated in **A** clef with ledger lines indicating the transition into the **α** range.

While traditional ledger lines extend the upper and lower range of staves by numerous lines and spaces, the number of ledger lines in TAB+ is limited to the number of pad rows. In the case of the Push, in **A** and **B** clefs, a maximum of four ledger lines above the tablature can be used for denoting upper extensions into the **α** and **β** range. In **α** and **β** clefs, a maximum of four ledger lines below the tablature denote lower extension into the **A** and **B** range. As demonstrated in Figure 5 (a), a two-octave scale can be notated with a clef change or (b) with ledger lines.

The TAB+ additionally includes indications for the parameters: *layout* (e.g., 4ths), *mode* (e.g., 'In Key'), *key* (e.g., C), and *scale* (e.g., major), ensuring that the notation is read correctly. The italic numbers above the staves specify fingerings and indicate transitions between pad rows (Fig. 4 and 5). Fingering suggestions for the first and last notes of the scale show the correct starting and ending positions.

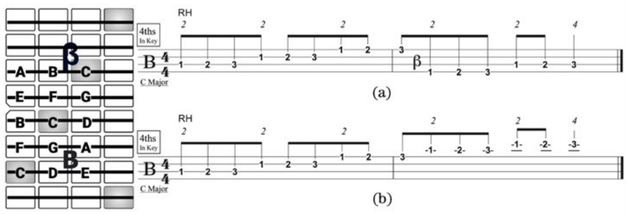


Figure 5. (a) An ascending two-octave C major scale (C3–C5) notated in **B** and **β** clefs for the right hand; (b) the same

scale notated in **B** clef with ledger lines indicating the transition into the **β** range.

The TAB+ system further introduces two hybrid clefs; **AB** and **αβ** (pronounced as AB clef and alpha-beta clef). The combination of the **A** and **B** clefs allows for the use of 32-pad layouts. The **AB** and **αβ** hybrid clefs are pertinent to the Sequent layout, for which the notation is more idiomatic using a 32-pad layout due to the arrangement of the pad rows in octaves. As observed in Figure 6, the tablature with hybrid clefs requires additional number indications (e.g., 5, 6, 7, and 8) for the Sequent layout.

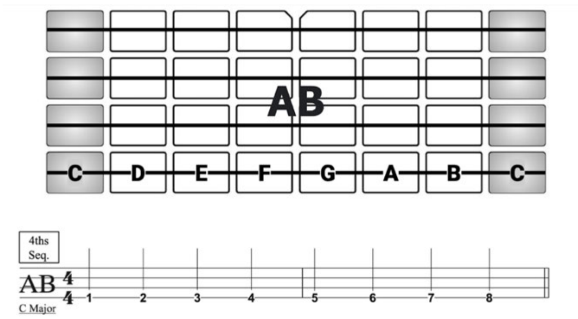


Figure 6. An ascending one-octave C major scale (C2–C3) in Sequent layout notated in **AB** clef.

In addition to facilitating notation in the Sequent layout, these hybrid clefs solve the challenges of notating musical material played *on* or *between* two nano layouts. For example, in a two-octave C major scale in thirds played with two hands, the left hand would be notated in **A**, and the right would be notated in **AB** and **αβ** clefs (Fig. 7). The **AB** and **αβ** clefs indicate that the right-hand moves across multiple nano layouts during the scale.

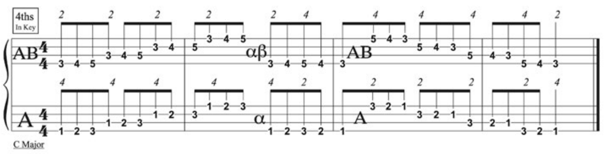


Figure 7. An ascending and descending two-octave C major scale in thirds played with two hands: right hand starting on E2 and left hand on C2.

The TAB+ system is transferable to other grid UIs. The number of lines in the tablature system can be adjusted depending on the number of rows of physical pads or graphical grids. Likewise, the numbering of the pads can be changed depending on the layout of the grid interface.

5.2. Adaptability of TAB+ in Different Push Layouts

The flexibility of the TAB+ is observed in its capacity to represent musical material in different modes, layouts, and directions. Figure 8 demonstrates how a two-octave C major scale (played by the left hand) can be notated in the 'In Key' mode for the 4ths, 3rds, and Sequent layouts. Figure 9 displays how the same scale can be notated in the 'Chromatic' mode for 4ths, 3rds, and Sequent layouts.

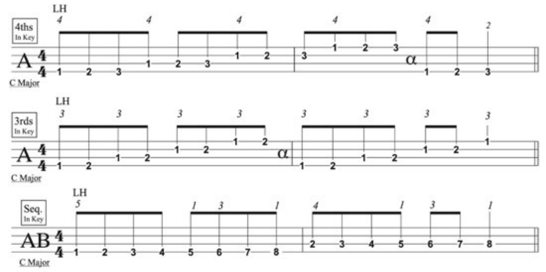


Figure 8. An ascending two-octave C major scale (C2–C4) notated for 'In Key' mode in 4ths, 3rds, and Sequent layouts.

The examples in Figure 8 highlight the flexibility of the TAB+ clefs, allowing the same musical material to be notated in three different ways, facilitating the readability and playability depending on various musical contexts. For example, guitarists accustomed to fourths tunings may prefer the 4ths layout, while pianists may find the Sequent layout more familiar. The TAB+ notation can be adapted to various artistic and pedagogical needs.

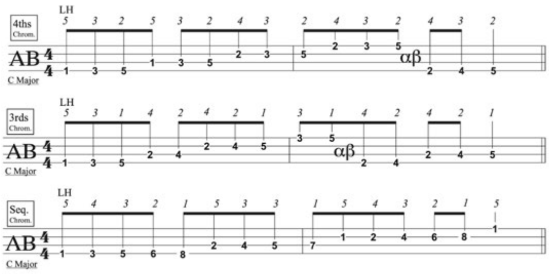


Figure 9. An Ascending two-octave C Major scale (C2–C4) notated for 'Chromatic' mode in 4ths, 3rds, and Sequent layouts.

The 'Chromatic' mode examples in Figure 9 further exemplify the functionality of hybrid clefs. These examples show that due to the extended playing range resulting from the addition of chromatic pitches, a 32-pad layout is required to simplify the notation and avoid excessive clef changes.

5.3. Harmonic Notation in TAB+

The TAB+ system provides the ability to represent chords. Figure 10 demonstrates root position diatonic triads in the key of C major played with two hands. In contrast to tablature systems used for fretted instruments, which typically use a single number per line to indicate the string and fret position, TAB+ utilizes multiple numbers on a single line to indicate that multiple pads need to be played within a specific row. Furthermore, Figure 10 demonstrates the application of ledger lines in chord notation. The first three diatonic triads notated in **β** clef with ledger lines below the tablature indicate that the notes of the chords are in both **β** and **B** nano layouts.

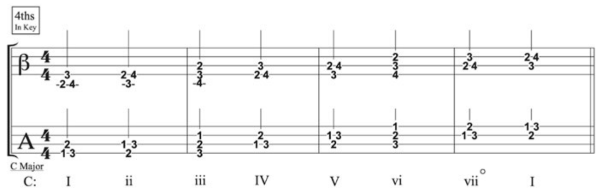


Figure 10. Root position diatonic triads in C major.

The examples thus far have demonstrated the use of TAB+ in the Vertical direction of the Push. The same notation principles are transferable to the Horizontal direction and other grid TUIs and GUIs with the capacity for changing the directional layout. Figure 11 illustrates the notation of a one-octave C Major (C2–C3) scale played by the left hand in the Horizontal direction.

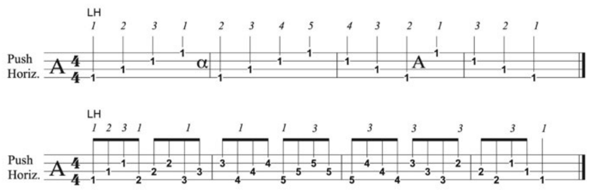


Figure 11. A one-octave C major scale (C2–C3) in 4ths Layout, 'In Key' mode, and Horizontal direction.

Indicating the direction parameter (i.e., vertical or horizontal) within a musical score is needed as it affects how the notation is read and realized in teaching, learning, and performance contexts. The TAB+ system incorporates the abbreviations used by Ableton in the Push display for vertical and horizontal, 'Vert.' and 'Horiz.' directions into the score (Fig. 12). This integration not only enhances the accuracy of musical notation but also streamlines musical communication.

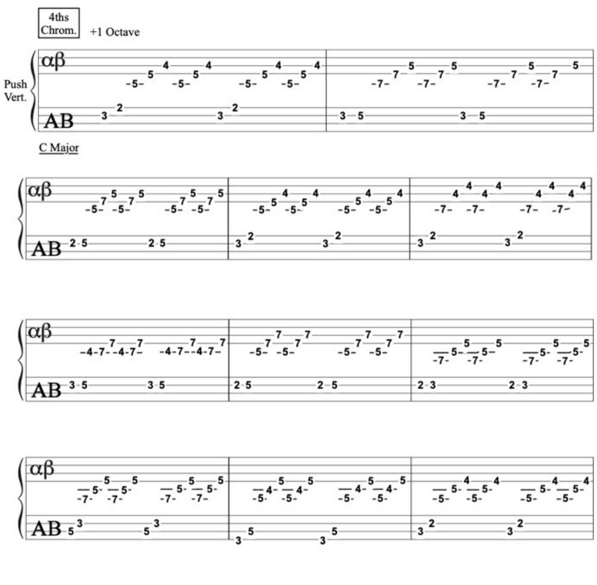


Figure 12. J. S. Bach, "Prelude No. 1 in C Major," from *The Well-Tempered Clavier, Book I*, measures 1–11, notated using the TAB+ system.

6. STAFF+: TRADITIONAL STAFF NOTATION WITH TAB+ CLEFS AND ROMAN NUMERAL MATRICES

6.1. Single-Line Notation in Staff+

The Staff+ subsystem of TABstaff+ is designed with the view that 'regardless of the many varied developments around the fundamental elements of music,' the traditional five-line staff notation offers immense 'flexibility' in terms of notation for electronic music [21]. As such, Staff+ expands on the traditional five-line staff notation system through the incorporation of the **A**, **B**, **α**, and **β** TAB+ clefs (introduced above) and Roman numeral matrices for labeling specific pads. Similar to TAB+, the alpha and beta clefs specify one of the four nano layouts. Unlike in TAB+, which indicates pads using tablature lines for pad rows and Arabic numerals (i.e., one to four and one to eight), the Staff+ system employs Roman numeral pairs (e.g., **(II, ii)**) for specifying individual pads. The uppercase Roman numeral indicates the location of the pad on the x-axis, and the lowercase numeral specifies the location of the pad on the y-axis (Fig. 13). The Staff+ system uses alpha-numeric indicators to clarify pitch and octave for notes on the conventional five-line staff and the corresponding pads on the grid-playing surface. These indicators combine either an alpha or beta clef with a Roman numeral pair (for instance, **B (II, iv)**).

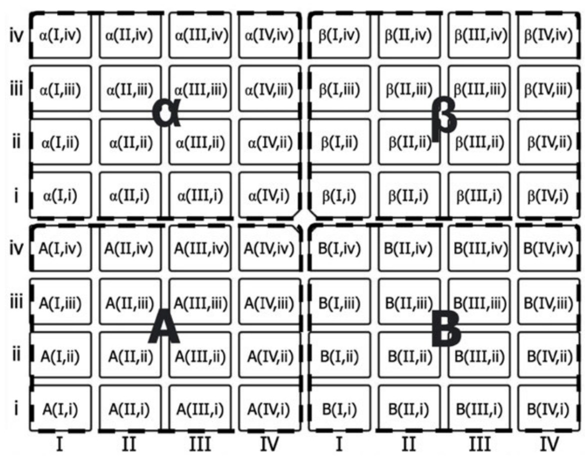


Figure 13. TAB+ clefs and Roman numeral matrices specifying 16-pad (4×4) nano layouts and individual pads.

Using alpha-numeric indicators, an ascending and descending two-octave C major scale (C4–C6) would be notated, as shown in Figure 14. The starting pitch (middle C) is labeled **B (II, iv)**. The **B** clef clarifies which nano layout notes are to be played in, and the Roman numeral pairs specify the pads corresponding to the notes. With this information, the user knows to position their right hand over the nano layout **B** in the bottom right of the Push and play the second pad (x-axis: **II**) in the fourth row (y-axis: **iv**). In Staff+, the alpha-numeric markers also serve as points of reference for transitions between different nano layouts. As seen in Figure 14, the **β (II, i)** specifies that with F4, the hand shifts to and remains in the upper right nano layout until the E4 in the last measure, where **B (IV, iv)** indicates the transition back to the bottom right nano layout. Similar to the TAB+ system, italic Arabic numerals are used for fingerings and transitions between pad rows (Fig. 14).

Figure 14. An ascending two-octave C major scale in Staff+ notation with TAB+ clefs and Roman numeral pairs indicating transitions between nano layouts.

For two-hand notation, Staff+ employs the conventional grand staff with treble and bass clefs. The use of traditional notation provides 'everything we need to capture the quintessential elements of music: pitch on the vertical, time on the horizontal, and anything else in the margins' [21]. Alpha-numeric indicators are used in both staves to designate notes and their corresponding pads. Figure 15 shows an example of a two-octave C major scale in sixths notated on a grand staff with treble and bass clefs, TAB+ clefs, and Roman numeral pairs, the amalgamation of which defines the Staff+ notation system.

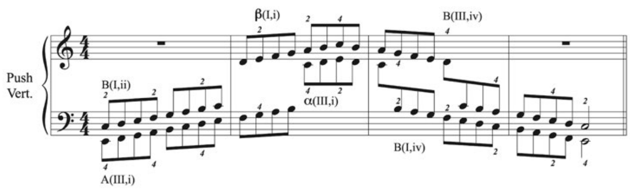


Figure 15. An ascending two-octave C major scale (in sixths) in Staff+ with TAB+ clefs and Roman numeral pairs indicating transitions between nano layouts.

In the above example, the right hand (treble clef) begins on pad **B (I, ii)**. As the scale ascends, **β (I, i)** is used to show the transition of the hand into the upper-right nano layout; as the scale descends, **B (III, iv)** designates the transition back to the lower-right nano layout. Similar indicators for the left hand (bass clef) are observed in Figure 15. As the left-hand traverses all four nano layouts during the scale, three TAB+ clefs (**A**, **α**, and **B**) are used to mark transitions between the lower half of the instrument (32-pads **(A, B)**) and the upper half of the instrument (32-pads **(α, β)**).

6.2. Harmonic Notation in Staff+

In Staff+, the TAB+ clefs with Roman numeral pairs can be used for notating chords and other harmonic material. As seen in Figure 16, diatonic triads in C major can be notated by indicating the lowest note of the chord. For example, a root position C major triad built on C4 would be notated using the **B** clef and the Roman numeral pair **(II, iv)**. By labeling the root position C major triad alpha-numerically with **B (II, iv)**, the register of the chord and its position on the instrument are made clear.

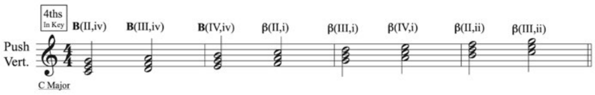


Figure 16. Root position diatonic triads in C major with TAB+ clefs specifying nano layouts and Roman numeral pairs indicating the position of the lowest note of each triad on the instrument.

Chord inversions are represented by specifying the position of the lowest note of a given chord. As seen in Figure 17, **B (IV, iv)** indicates a C major triad in the first inversion (lowest note E4), and **β (III, i)** specifies the second inversion (lowest note G4).

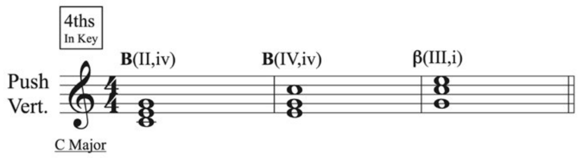


Figure 17. Staff+ labeling for inversions.

As demonstrated by the provided examples, Staff+ builds the traditional five-line staff notation by introducing TAB+ clefs and Roman numeral matrices. These elements are designed to aid in reading, learning, teaching, and performing music on the Push, other grid TUIs and GUIs.

7. CHARTS+: CHORD DIAGRAMS FOR THE PUSH AND GRID INTERFACES

7.1. Chord Diagrams for 'In Key' Mode

Charts+ is the third subsystem of TABstaff+; it utilizes 16, 32, and 64-pad diagrams for graphical representation of chords and playing actions such as the triggering of loops, sequences, formal sections, and other discrete musical elements [35]. The 16 and 32-pad diagrams are used primarily for chords in 'Note' mode on the Push, and 64-pad diagrams are used for specifying the activation and deactivation of loaded clips in 'Session' mode. Indicating chords using 64-pad diagrams is possible (Fig. 18); however, 16-pad and 32-pad diagrams offer the advantage of being easier to read, more space efficient in score formats, and resembling traditional chord diagrams used for instruments like guitar, ukulele, and others.

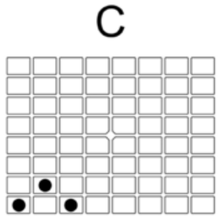


Figure 18. A Charts+ 64-pad diagram for a C major triad in root position.

In traditional chord charts, horizontal lines represent frets, and vertical lines represent strings, in Charts+, cell grids are used to represent pads. Similar to traditional chord charts, finger indications are given using black dots, which can be solid or contain numbers for finger indications. Light grey gird cells are used to show the tonic of a given scale (Fig. 19). Additionally, it is possible to specify scale, layout, and mode parameters within the charts.

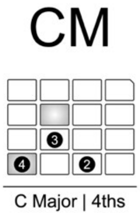


Figure 19. A Charts+ 16-pad diagram for a C major triad in root position with finger indications for the left hand.

The Charts+ system can be used in tandem with the Staff+ and TAB+ systems. Figure 20 shows how chord charts can be added above staves to specify finger placements corresponding to the notated triads.

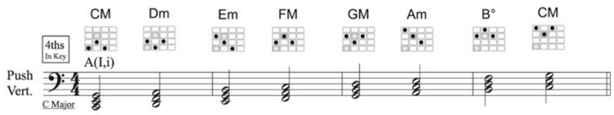


Figure 20. Root position diatonic triads in Staff+ with Charts+ diagrams.

Due to the different pad layout possibilities of the Push and similar grid instruments, there are multiple ways to play a given chord. For example, in the 4ths layout, a C major triad can be played with four different finger patterns. As observed in Figure 21, the C major triad can be played on a single row of pads (finger pattern one), two rows with two pads on the lower row and one pad on the upper row (finger pattern two), two rows, with one pad on the lower row and two pads an on the upper row (finger pattern three), and lastly, across three pad rows with a single pad per row (finger pattern four).

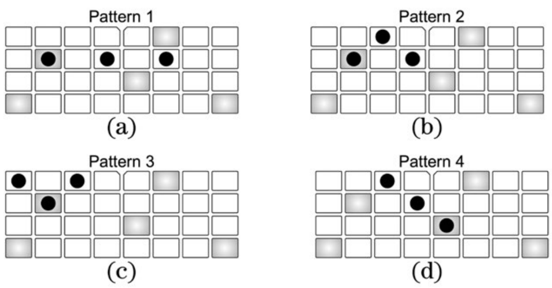


Figure 21. 32-pad Charts+ diagrams showing four finger patterns for a C major triad in root position.

Finger patterns two (b), three (c), and four (d) can be notated using 16 or 32-pad diagrams given that these lie on adjacent pad columns (Fig. 21). Finger pattern one (a) requires the use of a 32-pad chart, as it spans five pad columns, extending beyond the 4×4 grid of a 16-pad nano layout. Chords between two different 16-pad nano layouts can be represented graphically using a 32-pad diagram.

As observed in Figure 21, finger patterns one (a) and four (d) have two notes in the bottom left nano layout and one in the bottom right. This necessitates the use of a 32-pad diagram for graphical representation. The versatility of 16 and 32-pad chord diagrams can be observed in specifying extended chords and other complex harmonies. These diagrams enhance the comprehension of intricate chord structures, providing a valuable resource for teachers, students, and practicing musicians.

7.2. Chord Diagrams for ‘Chromatic’ Mode

The notation of chords in the ‘Chromatic’ mode on the Push requires 32-pad diagrams. As the addition of chromatic pitches makes finger patterns wider, two 16-pad nano layouts are needed for playing chords (Fig. 22). The Chart+ diagrams employ black and white grids to represent unlit and lit pads in ‘Chromatic’ mode.

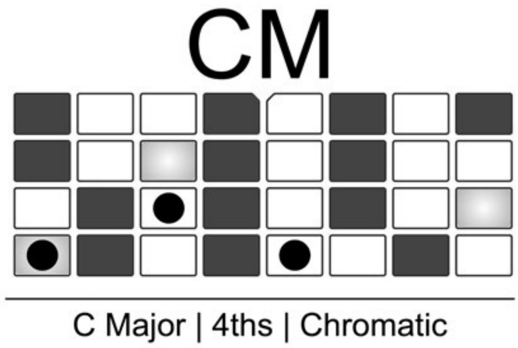


Figure 22. Charts+ diagram for a root position C major triad in the ‘Chromatic’ mode.

Similar to how a chord can be played with different finger patterns using the ‘In Key’ mode, the same holds in the ‘Chromatic’ mode. Figure 23 shows three finger patterns for playing the same F major triad. Each pattern offers unique flexibility depending on the musical or pedagogical context.

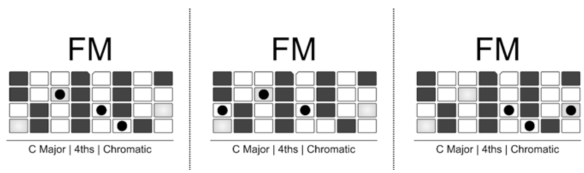


Figure 23. Three Charts+ diagrams showing different finger patterns for the same root position F major triad in the ‘Chromatic’ mode.

Charts+ offers a dynamic and versatile solution for artists and educators, seamlessly integrating notation with diverse TUIs and GUIs.

7.3. Diagrams for 'Session' View on the Push

The Charts+ system can be used in the 'Session' mode for indicating the triggering of clips containing musical fragments such as loops, song sections, and other elements at various levels of musical structure. The diagrams for the 'Session' mode use three grayscale colors for specifying empty, loaded, and active clips. Black pads indicate open clip slots, light grey pads specify loaded clips, and white pads show active clips. Black dots are used to indicate which pads should be pressed to launch clips (Fig. 24 a). On the Push, there are two main methods for stopping clips; the first is to press an empty slot within the track (column of pads) that contains the active clip (Fig. 24 b), and the second is to use the 'stop clip' button (Fig. 24 c). The '×' symbol is used to indicate which pads are to be pressed to deactivate clips.

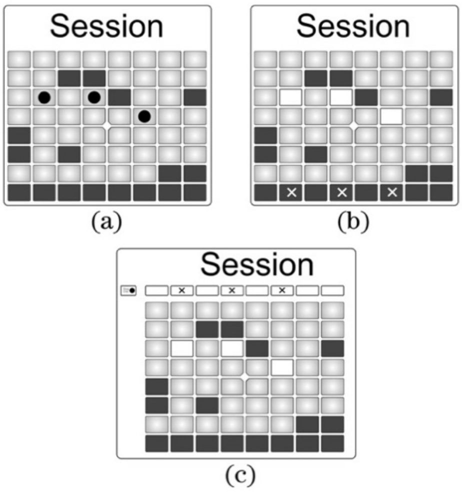


Figure 24. Charts+ diagrams for 'Session' view; (a) shows the triggering of clip slots, (b) specifies the stopping of clips using open slots, and (c) shows the stopping of clips using the 'stop clip' button and buttons below the display screen.

Instructions for deactivating clips using the second method are given in the chart (Fig. 24 c) by indicating pressing the 'stop clip' button in combination with the buttons below the display screen corresponding to the track(s) containing active clip(s). These examples illustrate the use of grid cell diagrams of various dimensions for the graphical representation of chords and other discrete musical structures. The diagrams presented above can be adapted to various musical grid interfaces and applied to a wide range of artistic and pedagogical contexts.

8. CONCLUSION

The TABstaff+ notation system presented in this paper is designed to provide a robust means of graphically communicating musical structures for musical grid interfaces. The paper emphasized the need for comprehensive notation systems that account for the unique attributes of these instruments. A new notation system, TABstaff+, was introduced as a versatile solution for this challenge. TABstaff+ encompasses three subsystems: TAB+, Staff+, and Charts+. These systems, while building on established notation conventions, consider the specialized layout of musical grid interfaces, allowing musicians and educators to communicate ideas more effectively. The TABstaff+ system leverages HCI and pre-digital practices of reading, playing, and teaching music with instruments and notation. The paper stressed the importance of music literacy, especially in a world where digital technologies are increasingly integrated into music education. TABstaff+ is presented as a means of enhancing pedagogical approaches, making it easier for students to learn, play, and comprehend music on grid-based instruments. The paper provided insights into the adaptability and flexibility of the TABstaff+ notation system across different layouts and modes, showcasing its applicability in various contexts. It highlighted the potential of TABstaff+ to empower musicians, educators, and learners by offering a systematic and user-friendly approach to musical notation that bridges the gap between traditional and digital music education. The implications of TABstaff+ resonate with broader discussions on HCI and the evolving relationship between technology, education, and the arts. As the digital landscape continues to shape the future of music, TABstaff+ aims to contribute to the development of postdigital artistic and pedagogical practices.

9. REFERENCES

[1] M. Pelleter, *Futurhythmaschinen: Drum-Machines und die Zukünfte auditiver Kulturen*. Hildesheim: Georg Olms, 2021. http://doi.org/10.18442/166.

[2] C. Roads, *Composing electronic music: a new aesthetic*. Oxford: Oxford University Press, 2015.

[3] M. J. Butler, *Playing with something that runs: technology, improvisation, and composition in DJ and laptop performance.* New York: Oxford University Press, 2014.

[4] P. Manning, *Electronic and computer music*, 4th ed. New York: Oxford University Press, 2013.

[5] R. Guérin, *MIDI power!*, 2nd ed., Boston, MA: Thomson Course Technology, 2006.

[6] V. Giles, "Scoring For Grid Controllers – Preliminary Considerations Of Action, Gesture, Functionality, And Limitations," Sydney: Australia, 2015.

[7] S. Giddings, "Electronic, Digital, and Virtual Instruments," in *Technology for Unleashing Creativity*, Oxford University Press, 2022, pp. 119–139. https://doi.org/10.1093/oso/9780197570739.003.0009.

[8] B. Rossmy, M. Rauh, and A. Wiethoff, "Towards User Interface Guidelines for Musical Grid Interfaces," in *NIME 2022*, The University of Auckland, New Zealand, 2022. https://doi.org/10.21428/92fbeb44.db84ecd0.

[9] C. Modder, "MIDI – A Comprehensive Guide | Part 1 – MIDI for Learning." [Online]. Available: https://coenmodder.com/midi-a-comprehensive-guide-to-using-midi-for-learning-music (accessed Nov. 23, 2023).

[10] W. I. Bauer, *Music learning today: digital pedagogy for creating, performing, and responding to music.* New York: Oxford University Press, 2014.

[11] W. I. Bauer, "The Acquisition of Musical Technological Pedagogical and Content Knowledge," *Journal of Music Teacher Education*, vol. 22, no. 2, pp. 51–64, Apr. 2013. https://doi.org/10.1177/1057083712457881.

[12] B. Rossmy and A. Wiethoff, "Musical Grid Interfaces: Past, Present, and Future Directions," in *NIME 2021*, Shanghai, China, 2021. https://doi.org/10.21428/92fbeb44.6a2451e6.

[13] A. Aciman, "Meet the unassuming drum machine that changed music forever," *Vox*. [Online]. Available: https://www.vox.com/culture/2018/4/16/16615352/akai-mpc-music-history-impact (accessed Feb. 29, 2024.)

[14] D. Barolsky, "Conference Report: 'Performance Studies Network Second International Conference' ", *Music Theory Online*, vol. 19, no. 2, 2013. [Online]. Available: https://mtosmt.org/issues/mto.13.19.2/mto.13.19.2.barolsky.html (accessed Nov. 23, 2023).

[15] A. Pras, C. Guastavino, and M. Lavoie, "The impact of technological advances on recording studio practices," *J Am Soc Inf Sci Tec*, vol. 64, no. 3, pp. 612–626, Mar. 2013. https://doi.org/10.1002/asi.22840.

[16] J. Knowles, "Performance recordivity: Studio music in a live context," *Creative Industries Faculty School of Media Entertainment Creative Arts*, Jan. 2012. [Online]. Available: https://www.academia.edu/74098783/Performance_recordivity_Studio_music_in_a_live_context (accessed Nov. 23, 2023).

[17] 'A Notation System for Ableton Push', Chris Perren. Accessed: Nov. 23, 2023. [Online]. Available: https://www.chrisperren.com/blog/2020/12/02/a-notation-system-for-ableton-push

[18] B. Folkes, 'Electronic Music Notation (PL Sep/Oct 1978)', *Polyphony*, no. Sep/Oct 1978 Vol. 4, Issue 2, pp. 9–14, Sep. 1978, Accessed: Nov. 23, 2023. [Online]. Available: https://www.muzines.co.uk/articles/electronic-music-notation/12327

[19] H. R. Gudmundsdottir, "Advances in music-reading research," *Music Education Research*, vol. 12, no. 4, pp. 331–338, 2010. https://doi.org/10.1080/14613808.2010.504809.

[20] M. Gaare, "Alternatives to Traditional Notation: Traditional notation has survived the test of time and is read by musicians around the world, but there are many musicians who desire, devise, or use alternative systems," *Music Educators Journal*, vol. 83, no. 5, pp. 17–23, 1997. https://doi.org/10.2307/3399003.

[21] C. Faia, "Notating electronics," *Musica/Tecnologia*, pp. 43–65, 2019. https://doi.org/10.13128/MUSIC_TEC-11162.

[22] "The power of music to change lives: a national plan for music education," *GOV.UK*. [Online]. Available: https://www.gov.uk/government/publications/the-power-of-music-to-change-lives-a-national-plan-for-music-education (accessed Nov. 23, 2023).

[23] "The future of education and skills: Education 2030. The future we want." OECD Publishing, 2018.

[24] M. Emo, "Ableton Live professional learning development for secondary school music teachers," *Journal of Music, Technology & Education*, vol. 14, no. 1, pp. 43–68, 2021. https://doi.org/10.1386/jmte_00033_1.

[25] M. Amico and L. Ludovico, "Kibo: A MIDI Controller with a Tangible User Interface for Music Education," in *Proc. of the 12th International Conference on Computer Supported Education*, 2020, pp. 613–619. https://doi.org/10.5220/0009805206130619.

[26] E. Frid, 'Accessible Digital Musical Instruments – A Review of Musical Interfaces in Inclusive Music Practice', *MTI*, vol. 3, no. 3, p. 57, Jul. 2019, doi: 10.3390/mti3030057.

[27] A. Hart, 'The Tablet as a Classroom Musical Instrument', *AJE*, vol. 5, no. 3, pp. 299–324, Aug. 2018, doi: 10.30958/aje.5-3-5.

[28] E. J. van Zoelen, "Klavar: From Tablature to Comprehensive Music Notation System," *Research & Issues in Music Education*, vol. 17, no. 1, Feb. 2023, [Online]. Available: https://commons.lib.jmu.edu/rime/vol17/iss1/5

[29] G. McPherson (ed.), *The Child as Musician: A handbook of musical development*. New York: Oxford University Press, 2006. https://doi.org/10.1093/acprof:oso/9780198530329.001.0001.

[30] R. Orlando and C. Speelman, "The Effect of Music Teaching Method on Music Reading Skills and Music Participation: An Online Study," *Research outputs 2012*, Jan. 2012, [Online]. Available: https://ro.ecu.edu.au/ecuworks2012/846

[31] M. B. Beckerman, P. A. Boghossian, and Open book Publishers, Eds., *Classical Music: contemporary perspectives and challenges*. Cambridge, UK: OpenBook Publishers, 2021.

[32] D. J. Hargreaves, Ed., *Children and the arts*. Milton Keynes [England] ; Philadelphia: Open University Press, 1989.

[33] "Ableton Push 3 Manual." [Online]. Available: https://www.ableton.com/en/push/manual (accessed Nov. 23, 2023).

[34] "Using Push 2. Ableton Reference Manual." [Online]. Available: https://www.ableton.com/en/manual/using-push-2 (accessed Nov. 23, 2023).

[35] W. Kuhn and E. Hein, "Ableton Live and Push," in *Electronic Music School*, Oxford University Press, 2021, pp. 40–45. https://doi.org/10.1093/oso/9780190076634.003.0005.

MOTION CAPTURE DATA AS MACHINE-READABLE NOTATION TO CAPTURE MUSICAL INTERPRETATION: EXPERIMENTING WITH MOVEMENT SONIFICATION AND SYNTHESIS

Julien Mercier
Media Engineering Institute (MEI),
School of Engineering
and Management Vaud, HES-SO
julien.mercier@hes-so.ch

Irini Kalaitzidi
Department of Computational Arts,
Goldsmiths University of London
eri.kalaitzidi@gmail.com

ABSTRACT

Musical notation can be described as an abstract language that composers use so that performers may interprets a score. Such notation comes before interpretation, is reproducible, and although it contains hints targeted at the performer on how to interpret, the actual performance is uniquely situated in time and space. The only way to record and re-experience a sound performance is to use microphones, which transform acoustic waves into an electrical signal and usually lose at least some spatial dimension in the process. This may be hindering in the field of experimental music, where physical limits of sound material may be put to the test. In this short paper, we discuss how motion capture could be an alternative to, or an expansion of the acoustic recording of a performance involving movement. By recording the performer's movements, some of the dimensions that make their interpretation singular (i.e., character, accentuation, phrasing, and nuance) are retained. A method capturing sound through movement may be interesting in the context of sound synthesis with deep learning and hold potential advantages over current methods using MIDI or acoustic, which either lack dimensions or are very sensitive to noisy data. We briefly discuss rational, practical, and theoretical foundations for the development of potentially innovative outputs.

1. INTRODUCTION

Traditionally, musical notation serves as a common, abstract language for musical composition that doesn't specify all aspects of musical performance. In a second time, various performers must seize this coded language and make various choices during the interpretation phase. It's a vector between different humans involved in the creation and interpretation of sound. The actual performance, which is situated in time and space, cannot be notated in such an abstract way; it can merely be recorded as an acoustic output: The acoustic waves are converted into an electrical signal by a transducer before it is sampled in order to be encoded digitally. During live performances, many of the spatial qualities of sound are usually lost in the process, because microphones typically capture sound stemming from various sources and merge them. In a studio setting, this is avoided by recording every instrument separately. Altogether, the acoustic recording process is currently the only format through which a performance and the interpretation it embodies are captured in order to be replicated afterwards.

2. MOVEMENT SONIFICATION WITH MOTION CAPTURE

Current advances in computer vision algorithms make it possible to capture and record the motion of a performer (whether hands, face, or whole body) in real time with great levels of accuracy, thus capturing different dimensions than those recorded by a microphone. By doing so, it is possible to capture the source impulse that yields the distinctiveness of a performance rather than its mere output, as is the case with acoustic recording. As early as 1938, Alexander Truslit's [1] used a rudimentary motion capture system to elaborate his theory on the gestural quality of musical interpretations. He researched the relationship between the motion of music and the perceptual processes of the audience. Wöllner [2] investigated Truslit's hypothesis by recording and comparing free movements and instructed movements and asking participants to determine which sound they had produced in a self-other recognition task. Motion recording has the potential to capture some of the subtle and sensory dimensions that make it an interpretation singular: character (the general hue given to the expression of a piece), accentuation (Agogik), phrasing, and nuance. It results in a type of numerical notation that can only be read by machines in order to re-enact a performance at a later stage. As such, capturing the performer's movement represents an interesting alternative method to acoustical recording for the technical repeatability of a performance. In fact, it may prove a particularly fitting method in some cases. It can be used for the automatic notation of improvised music without the need for post-hoc transcription, which usually doesn't do justice to the creativity of the exercise. If the instruments used by the performers are digital, then every unit of a similar model is identical, and the parameters used can be reproduced.

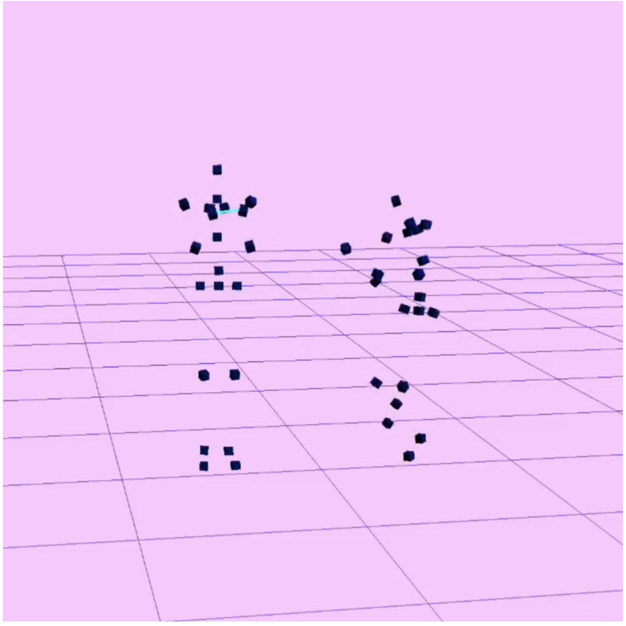


Figure 1. Once recorded by a MaxMSP program, the performers' movements can be played again and sonified with the same synthesizer. Provided a similar setup, there is no difference between the sounds produced during the original performance or its re-enactment.

The conditions under which the performer's movements are turned into sound (including the spatial setup) can be faithfully reproduced across different places and times. In the case of experimental music, the spatial qualities of the sound are critical to ensuring proper perception, and acoustic recording systems can sometimes reach their physical limits. In specific cases where digital musical instruments can be parameterized in absolutely identical ways, recording a performance through motion capture may be equivalent to "higher-dimensional MIDI data" and enable a more naturalistic reproduction of the sound produced by the performance. Each instrument has its own specificity, and an approach based on motion recording may prove interesting in various situations.

3. USE CASE: DANCE IMPROVISATION SONIFICATION

The use of motion capture in combination with dance and movement sonification is the object of multi-disciplinary research and encounters a variety of applications in data sonification and gesture interfaces, and it is observed through frameworks such as affective computing [3]. Movement sonification fosters new perspectives for practice-based artistic research [4], and it is an effective setup to achieve emotion-to-sound translations and is investigated by composers and sound designers to communicate affective information [5]. Movement sonification is often used to expand motion natural acoustics, and there are many ways to transform movement data into sound [6]. Real-time movement sonification has also been used in the field of gestural rehabilitation [7].

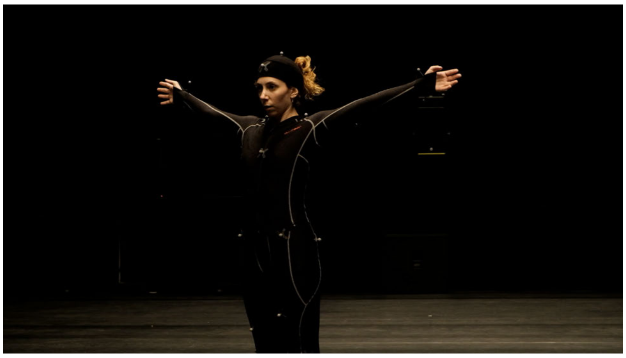


Figure 2. Screen capture of the performer wearing the motion capture suit and improvising with real-time auditory feedback. A video of the setup is visible at the following link: vimeo.com/783283941

We experimented with motion capture and movement sonification during an artistic research residency at the Institute for Computer Music and Sound Technology in 2020. We created a series of simple synthesizers with MaxMSP controlled by body motion through the use of a large motion capture system. The distance between the performer's hands and feet controlled simple sine-wave oscillators, while their X and Y position in the room controlled additional granular synthesizers. We recorded sessions during which a dancer performed various improvised choreographies intuitively, guided only by real-time sonification feedback. This type of auditory feedback has a positive influence on perception and motor movement, such as weight distribution, joint angles, and jumps are recognized through sound. By practicing, someone can determine through sound whether a complex movement was correctly executed. [8]. Our synthesizers were designed to turn specific types of movement into simple sounds. As such, re-enacting the performance by re-playing the motion data sequence produces rigorously similar outputs as the actual performance did, if the track is played on similar hardware, somehow inverting the notational process: from the performer to the composer. The motion capture data kept idiosyncrasies, which in this context are an operationalization of the concept of interpretation. Under different circumstances, the method could possibly be used to capture the hand movements of a pianist [9] and encapsulate some usually hidden dimensions of interpretation in this particular type of machine notation. However, the gestures we experimented with have a particular type of expressiveness, and the method might be only suited to the context of emerging virtually-controlled music performance [10].

4. DISCUSSION

We ask ourselves whether data captured while playing an instrument based on body movements may open interesting perspectives in the field of sound synthesis with deep learning. To our knowledge, movement sonification and movement capture have not been used to train deep-learning models for sound synthesis. Current methods either rely on the use of musical notation data (i.e., sequences of notes) or on the use of sampled audio data formats. As mentioned earlier, both have limitations:

- In the case of synthesis of musical notation data (i.e. MIDI), new sequences are generated by using recurrent neural networks (i.e., LSTM), which are good at memorizing past states of sequential data. Outputs take the form of new sheet music. The advantages is that it produces "clean" data, while the main drawback is that the data contains few dimensions and doesn't encapsulate any of an interpretation's dimensions. Synthesis based on musical notation has famously been used by smartphone company Huawei to "complete" Schubert's Symphony No. 8, although the result was deemed unsatisfactory [11].
- In the case of audio file data synthesis, it is usually transformed into graphical data first (i.e., Mel spectrograms) and processed with computer vision algorithms which synthesize new spectrograms. These are then translated back into sound samples, and the result is usually extremely sensitive to the quality of the recording. The advantage of this approach is that it may encapsulate human interpretation dimensions, but the drawback is that it produces very noisy outputs. Unless training on uncompressed data of sounds recorded in an anechoic chamber with close-to-perfect conditions, results are still far from satisfactory.

We formulate the hypothesis–without being able to answer it as of now–that using motion capture data whose recorded movements produced consistent sounds could be an interesting approach to training a model capable of generating new sequences, using the same techniques as musical notation synthesis (LSTM or Transformer architectures), while encapsulating dimensions that are related to the interpretation as represented in the training data. It would benefits from the advantages of both mentioned methods: clean, notation-based outputs and the encapsulation of interpretation dimensions. It would also avoid their pitfalls: unlike most notation systems, motion capture data would retain multiple dimensions related to the singularity of an interpretation, while remaining immune to noise in the training data the way sound synthesis through spectrogram is. It may also be that some other unsuspected sensitive dimensions would be lost in translation. This paper makes no claim as to the actual results yielded by this approach. Rather, it proposes to lay down and discuss rational, practical and theoretical foundations for future development of a system aimed at innovative outputs in the field of sound synthesis with deep learning.

Acknowledgments

The authors would like to thank the ZHdK's Institute for Computer Music and Sound Technology for hosting our artistic research: Eric Larrieux, Tobias Gerber, Martin Fröhlich, and Kristina Jungic.

5. REFERENCES

[1] H. Brandner and M. Haverkamp, *Alexander Truslit: Gestaltung und Bewegung in der Musik.* Reprint of the original edition 1938. Augsburg: Wißner, 2015.

[2] J. Hohagen and C. Wöllner, "Movement Sonification of Musical Gestures: Investigating Perceptual Processes Underlying Musical Performance Movements," in *Proc. of the 13th Sound & Music Computing Conference (SMC)*, Hamburg, 2016.

[3] S. Landry, J. Ryan, and M. Jeon, "Design Issues and Considerations for Dance-Based Sonification," in *Proc. of the 20th International Conference on Auditory Display (ICAD2014)*, New York, 2014.

[4] A. Giomi, "Somatic Sonification in Dance Performances. From the Artistic to the Perceptual and Back," in *Proc. of the 7th International Conference on Movement and Computing (MOCO '20)*, New York, 2020, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/10.1145/3401956.3404226

[5] S. Landry and M. Jeon, "Interactive Sonification Strategies for the Motion and Emotion of Dance Performances," *Journal on Multimodal User Interfaces*, vol. 14, pp. 167–186, 2019.

[6] A. Effenberg, "Movement Sonification: Effects on Perception and Action," *IEEE MultiMedia*, vol. 12, no. 2, pp. 53–59, 2005.

[7] T. H. Nown, P. Upadhyay, A. Kerr, I. Andonovic, C. Tachtatzis, and M. A. Grealy, "A Mapping Review of Real-Time Movement Sonification Systems for Movement Rehabilitation," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 672–686, 2023.

[8] T. Grosshauser, B. Bläsing, C. Spieth, and T. Hermann, "Wearable Sensor-Based Real-Time Sonification of Motion and Foot Pressure in Dance Teaching and Training," *Journal of the Audio Engineering Society*, vol. 60, no. 7/8, pp. 580–589, 2012. [Online]. Available: https://www.aes.org/e-lib/online/browse.cfm?elib=16369

[9] W. Goebl and C. Palmer, "Temporal Control and Hand Movement Efficiency in Skilled Music Performance," *PLOS ONE*, vol. 8, no. 1, 2013. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0050901

[10] A. Bouënard, M. M. Wanderley, S. Gibet, and F. Marandola, "Virtual Gesture Control and Synthesis of Music Performances: Qualitative Evaluation of Synthesized Timpani Exercises," *Computer Music Journal*, vol. 35, no. 3, pp. 57–72, 2011. [Online]. Available: https://doi.org/10.1162/COMJ_a_00069

[11] D. Szelogowski, "Generative Deep Learning for Virtuosic Classical Music: Generative Adversarial Networks as Renowned Composers," 2021. [Online]. Available: https://arxiv.org/abs/2101.00169

DIVINE RATIONALITY: COMPUTATIONAL TRANSCRIPTION OF CHANTS AND PARAMETRIC STRUCTURING FOR COMPOSITIONS FOR VOICES AND INSTRUMENTS

Dániel Péter Biró
The Grieg Academy – Department of Music
Faculty of Fine Art, Music and Design
University of Bergen
dpbiro@uib.no

Peter van Kranenburg
Department of Information and Computing Sciences
Utrecht University
p.vankranenburg@uu.nl

ABSTRACT

Our ongoing research into computational ethnomusicology, undertaken first in the project *Computational Ethnomusicology* and continuing in the projects *Tunes and Tales* and *Sounding Philosophy,* has allowed for a better understanding of the roles of scales, melodic contour and tuning in Jewish, Islamic and Christian chant traditions. While this research has also given way to new modes of notating chant cultures via computational means, it has also had a profound effect on the creation of what Dániel Péter Biró has termed historicized composition, a way of creating compositions that respond to, in this case, the historical development of chant traditions. To this end, the methodologies used for computational transcription of these chant traditions have been incorporated in compositional sketches and in innovative notational frameworks for new vocal music compositions. The following paper explores the methodology of this research, its technical properties and resulting compositional output, presenting examples from vocal and instrumental compositions completed between 2014 and 2022. Finally, we propose future work in these interdisciplinary research areas.

INTRODUCTION

Since 2011 we have been working collaboratively to analyze various chant traditions with computational means. This research was an outgrowth from Dániel Péter Biró's PhD research into Hungarian, Jewish and Christian chant traditions [1] as well as Peter van Kranenburg's PhD research into computational analysis of Dutch folksongs [2]. The initial fieldwork in the Netherlands, done as part of the *Tunes and Tales* research project involved recording members of Jewish and Islamic communities in the country. Each recording was then converted to a sequence of frequency values using the YIN pitch extraction algorithm [3]. The analysis of scales, melodic contours and tuning of these examples inspired a new kind of compositional methodology, allowing for the structuring of musical parameters such as pitch, timbre, tuning and instrumentation. Translating the various semiotic and phonetical structures found within the studied chant traditions, new notational frameworks were created.

1. COMPUTATIONAL ANALYSIS OF CHANT REPERTOIRES

The use of models deriving from various chant traditions has a long history. In the 20th century, the work of Olivier Messiaen long existed as a model of compositional integration of chant from various religious and world traditions. While Jewish and Islamic recitation traditions have been integrated within historical, twentieth and twenty-first century compositions by, among others, Josef Tal, Tristan Murail, Kaija Saariaho, Jonathan Harvey, Hans Zender, Samir Odeh-Tamimi, Katia Makdissi-Warren and José Maria Sanchez Verd, the integration of phonetic and semiotic structures found in these traditions has often been a secondary objective. The work of Béla Bartók (1881–1945) still exists as paramount in terms of being a model for connecting ethnomusicological research and musical creation.

In terms of integrating elements from Jewish, Islamic Christian chant traditions, the specific nature and functionality of each tradition had to be considered.

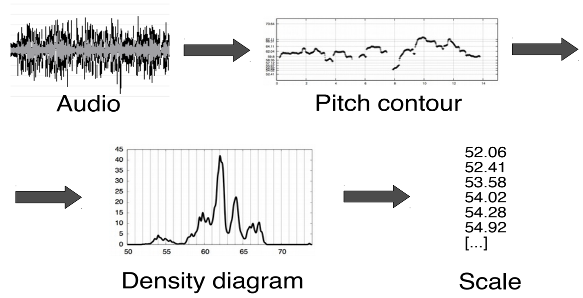
The accents for Torah cantillation are called *ta'amei hamikra,* which translates into "the taste" or "the meaning" of "the reading." These melodic signs, developed between the 7th and 9th centuries by the Masorete rabbis, were concurrently inscribed along with the vowels of the Hebrew letters in order to ensure accuracy in future Torah reading, thereby altering the previous mode of oral transmission. Within its ritual function, the Torah is not only sung but "read," as both syntactical and semiotic meaning is determined by the melody, which also functions to enable correct pronunciation. The ritual of "reading" exists to interpret the sacred Hebrew text, the textual sanctification enabled through the transmission and remembrance of a given Torah trope melody [1].

The word 'Qur'an' refers to the root q-r, and can be understood as "the recited."[4, p. 27]. [1] The performance framework for Qur'an recitation is governed by the rules of recitation. Here the hierarchy of spoken syntax, expression and pronunciation play a major role in determining the rules of *Tajwīd* and *tartīl* [5, p. 21].[2] The resulting melodic phrases, performed as "recitation" are determined by both religious and larger musical cultural contexts. In the context of "correct" Qur'an recitation contexts, improvisation and repetition exist in conjunction. The first Christian plainchant neumes appeared in the 9th century, evolving from a logogenic culture that was based on textual memorization, the memorized singing of chants being central to the preservation of a tradition that developed over centuries [6, p. 445].[3] Like with the Jewish *te'amim*, text existed as a precondition for the development of melody, the words of the chants being notated a century before the first neumes. Christian plainchant notation became a device that not only functioned to reconstruct melody, an aide-mémoire, but also to musically interpret textual pronunciation, syntax and meaning [1].

Within the framework of our research, we have recorded and analyzed examples of Qur'an recitation and Torah trope in various countries including the Netherlands, Tunisia, Israel, Germany and elsewhere [7], [8]. The recordings were put into a database and studied according to a variety of criteria including tuning, melodic contour, melodic range and scale employment (Example 1).[4] This research was also accompanied by interviews with the reciters about their background, including family history, their religious education and general questions about their ritual practice. This ethnographic knowledge informed the analysis methods and the integration and translation of elements from these traditions within the given compositions.

2. HISTORICIZED COMPOSITION INFORMED BY COMPUTATIONAL ANALYSIS OF CHANT

Over the last 30 years, Dániel Péter Biró has composed a series of works informed by Hungarian, Jewish, Christian and Islamic chant traditions. Since 2009, this compositional work has been increasingly informed by computational analysis of these chant traditions first undertaken in the context of the project *Computational Ethnomusicology* and carried forward in the projects *Tunes and Tales* and *Sounding Philosophy* [7], [8], [9], [10].



Example 1. Computational analysis methods allow one to analyze the scales, tuning and melodic contours of given recitation performance. Pitch is represented as MIDI note numbers, with the deviation in cents as decimals.

Within these projects, we recorded a given reciter reciting the same recitation over a series of years to test how memory plays a role in terms of melodic stability. Computational analysis of the resulting audio recordings was undertaken to find the most prevalent pitches within a given recitation.

In this research, we have set out to compare the performance practice of Torah trope recitation as it is done in a variety of cultural contexts. The computational method that we developed delineates the main pitches of a given scale from secondary "ornamental" pitches. By doing so, we presented a hierarchy of scale degrees, thereby showing how surrounding "ornamental" pitches structurally interact with the main "skeletal" notes of the scale within and across Torah recitation traditions.

In terms of studying *Qur'an* recitation, we were able to investigate the practice of *maqamat* (a set of pitches for melodic performance found in Qur'an recitation) in terms of phrase analysis to demonstrate the most prevalent pitches within a given performance in terms of foreground, middle-ground and background frequency analysis. While these pitches relate roughly to *maqamat* traditions of instrumental music, the melodic entities within Qur'an recitation do not adhere strictly to these [11, p. 9].[5]

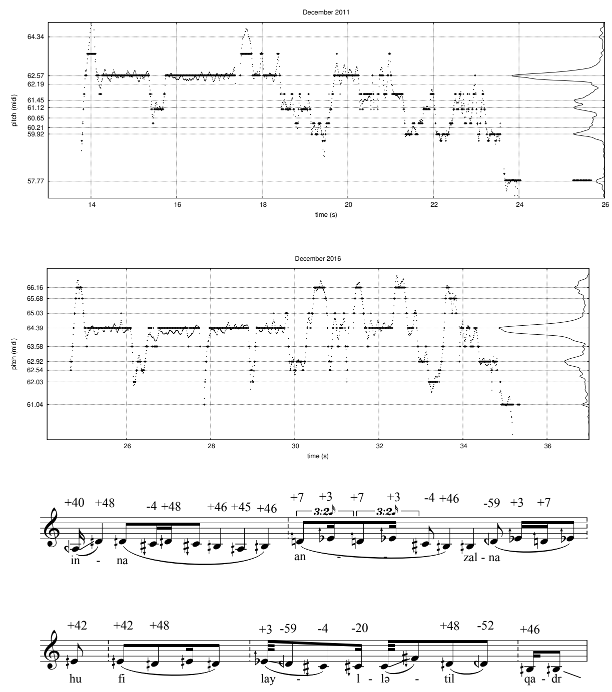
[1] "Like the Hebrew *miqra'* the primary name 'Qur'an' derives from the root q-r, i.e., 'reading': the visual implication of text is not implied with this root. Rather the concepts 'pronounce, calling, reciting' are expressed with the word, so that an adequate translation of Qur'an (Qur'ãn) could be 'the recited'" (Translation from German by Dániel Péter Biró).

[2] "*Tajwīd* [is] the system of rules regulating the correct oral rendition of the Qur'an. The importance of *Tajwīd* to any study of the Qur'an cannot be overestimated: *Tajwīd,* preserves the nature of a revelation whose meaning is expressed as much as by its sound as by it content and expression, and guards it from distortion by a comprehensive set of regulations which govern many of the parameters of the sound production, such as duration of syllable, vocal timbre and pronunciation."

[3] "Through the studies of Solange Corbin it has become evident that the neumes are of Carolingians origin. They were developed in France in the ninth century [...] Perhaps neumes were developed and used at first for theoretical demonstrations, and only occasionally employed to notate a particular melody or to give musical explanations here or there in a particular manuscript."

[4] Each recording has been segmented in terms of syntactical units (phrases) and analysis has also been based on audio segments corresponding to individual words of a given Qur'an *sura* or Torah *parasha.* Each recording has been converted to a sequence of frequency values using the YIN pitch extraction algorithm [3] by estimating the fundamental frequency in a series of overlapping time-windows of 40ms, with a hopsize of 10ms. The frequency sequences have been converted to sequences of real-valued MIDI pitches with a precision of approximately 1 cent (which is 1/100 of an equally tempered semitone, corresponding to a frequency difference of about 0.06%). A MIDI-value of 60 corresponds with the c', 61 with equal tempered c#', 62 with d', and so on. A value of e.g., 60.23 would correspond to a pitch that is 23 cents higher than c'.

[5] "Although Qur'anic recitation does not adhere strictly to the modal (maqam) practice of the secular music, and although members of the culture maintain a notion of rigid boundaries separating Qur'anic chant from all the other sound arts, Qur'anic recitation does conform to many of the theoretical aspects of Arabian music. It employs many of the interval combinations (trichordal, tetrachordalor, pentachordal) that identify apjns (s. jins) on which the secular music is based – e.g., bayyatl, ijaz,



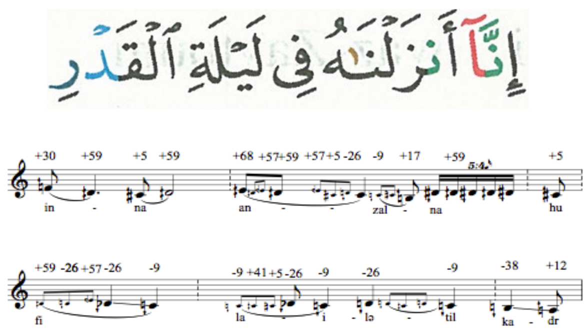
Example 2. Pitch analysis of recordings of recitation Sura *Al Qadr* performed in 2011 (top) and 2016 (bottom) with accompanying transcription as performed by a member of the Al Hikma Mosque in the Hague, Netherlands.

In our analysis, we have often compared renderings of the same reading of the same passage, done several years apart by a given reciter. Comparing the details of these readings reveals patterns of stability and – to a lesser extent – variation over time. Example 2 shows the pitch analysis of two readings of one ayah of Surah *Al-Qadr* as recited by the same reciter in the Netherlands, more than four years apart. We observe that the contour of the phrase shows stability, while the actual pitches, and consequently the scales, show different characteristics.

Such computational analysis provided scale tones for the transcription of these recitations with exact cent values, as shown in Example 3. Such methods of computational transcription, allowed for the basis of scales employed in the works composed in this period.

3. COMPUTATIONAL TRANSCRIPTION OF CHANT AND COMPOSITIONAL TRANSLATION

In this project, the computational analysis allowed for a new kind of compositional sketch, which became the basis for ensuing compositional procedures to structure sonority and form. Simultaneously, such computational models



Example 3. Transcription of Sura *Al-Qadr:* Recited in 2011 by a member of the Indonesian Muslim Community in The Hague.

needed to be interpreted in terms of their cultural context and accompanying recitation praxis. This required additional types of notation that would take the practice of vocal music beyond notational frameworks where the parameters of pitch and rhythm remain primary.

The composition *Al Ken Kara (That Is Why It Was Called)* (2013-2014) dealt with the Hebrew Bible text of the Tower of Babel (Genesis 11:1–9) [9]. The piece not only attempted to create a musical analogy to the breakdown of a unified, comprehensible language but also to show how various religious traditions created diverging types of phonetic, syntactical and notational contexts. Employing thirty-six languages, the composition integrates melodies from Ashkenazi Torah trope, Tunisian Qur'an recitation and plainchant. Ahead of the completion of the composition, the composer was able to research Jewish and Islamic chant practices in Tunisia and these became integrated into the work.

Toward the end of the composition, fragments of the Qur'an sura *Ghafir,* which in the composition is recited by the tenor, are incorporated into the now established polyphonic framework of the composition. The fragmented citations of Qur'an recitation, derived from Sura *Ghafir,* which tells of a tower that is, reminiscent to that of Babel:

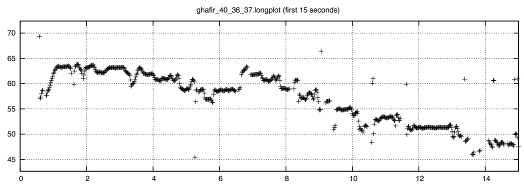
وَقَالَ فِرْعَوْنُ يَٰهَٰمَٰنُ ٱبْنِ لِى صَرْحًا
لَّعَلِّىٓ أَبْلُغُ ٱلْأَسْبَٰبَ ٣٦

Pharaoh said: "O Haman! Build me a lofty palace, that I may attain the ways and means [12, p. 738].[6]

Ahead of writing the composition, a Tunisian rendition of sura *Ghafir* was analyzed via computational means. Example 4 shows the computational contour transcription of

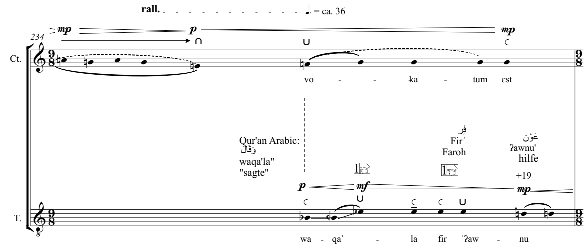
kurd, rast, nahawand, saba, slkah, etc. It similarly evidences a predominance of serial treatment of individual ajnas rather than utilization of the whole modal scale in a single phrase. Cantillation of the scripture is punctuated, like secular improvisations, by returns to the tone (qarar) of resolution in the jins. It is marked by transpositions and modulations internal or external to the phrase, which are also characteristic of the secular genres. Some reciters are knowledgeable about maqam practice and the theory of music. Others have no formal exposure to music theory and only conform to its rules in such measure as their ears and listening experience have trained them."

[6] "Muslim exegesis recognizes the presence of Haman alongside Pharaoh in building the tower of Babel in Q. 28:36/42 and Q. 40:36-7." Also see [13, p. 40].

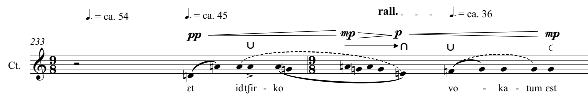


Example 4. Contour analysis of sura *Ghafir*.

Example 5. Transcription of Qur'an sura *Ghafir* as recited in Tunisia.



Example 6: Presentation of section of *Al Ken Kara (That Is Why It Was Called)* incorporating text and melody from Qur'an sura *Ghafir* as recited in Tunisia with scale pitches derived from computational analysis, the rhythmic profile interacting with the line of the countertenor, which is derived from a 14th century Czech plainchant [14].



Example 7. Agogic notation used in *Al Ken Kara (That Is Why It Was Called)* in conjunction with plainchant melody sung by countertenor.

the first 15 seconds. This analysis allowed for a transcription based on 10 scale tones, displayed in Example 5.

Within this composition for seven voices, this sura is cited, its scale becoming the basis for the melodic line of the tenor. Here *tajwīd* and *tartīl,* the rules governing the correct oral rendition of the Qur'an, are integrated into the notation [9], [5]. The signs above the notes indicate temporal elongation, constriction, acceleration and de-acceleration as well as mouth placement for the correct Arabic pronunciation. These elongations were based on the computational analysis, which set out to define how processes of elongation (Arabic: *madd)* could be transcribed within a cross-cultural music composition [7, p. 75].[7]

While the pitches for this transcription were derived from the computational analysis, a suitable notation needed to be found for the complex rules of pronunciation and elongation *(madd)* as found in Qur'an recitation [11, p. 10][8], [7], [9]. Here the signs above the notes served to indicate where a tone was shorter ᑕ or longer ᑌ which allowed for a certain flexibility in terms of interpreting the temporal structures present in our analysis of the sura *Ghafir* recitation. This notation to indicate agogic changes is also employed for the transcription of the Czech plainchant based on the text of the tower of Babel "Et idcirco vocatum est Babel nomen loci illius quia ibi divisum est labium universae terrae" [15, p. 16].[9]

4. VOCAL NOTATION INFORMED BY QUR'AN RECITATION

Within these compositions, *tajwīd*[10] and *tartīl*,[11] the rules and practice governing the correct oral rendition of the

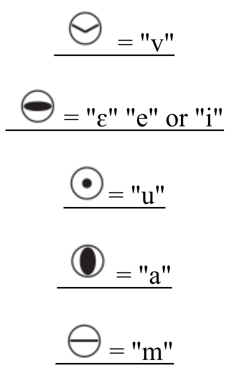
[7] "In our computational analysis we have looked for salient pitches within a recording, within each phrase and within words with syllable elongations *(madd).* Comparing the pitches employed in sections of textual elongation with those employed throughout a given recitation, we have found that the rules of *tajwīd* display a profound influence on creating and stabilizing salient pitches. In addition, through comparative analysis of the same *sura* recited by the same person, we are able to show scale relationships and recurring patterns of final selections of tones within these sections of elongation."

[8] "Durations of tones and rhythmic motifs are strongly affected by the rules of pronunciation set down in the manuals on *tajwid.* Those rules prescribe determined durational relationships between the short vowels or barakat (the fathah, dammah, and kasrah) and the long vowels (i.e., the letters alif, waw and Tajwld also determines the extension or madd of the long vowels according to their place in the word, their combination with certain other letters of the Arabic alphabet, and their use with unvowelled consonants (i.e., with sukon) or doubled consonants (tashdld. 7) These rules insure that the difference between the short and long syllables does not exceed the ratio of 1 to 6 (e.g., the difference between a 16th note and a dotted quarter). Often the actual differences are much less. Many of the prohibited practices regarding Qur'anic chant, which have been repeated and reemphasized in each successive century of Islamic history, have been restrictions against vocal practices that exaggerated durational contrast. Among these condemned practices are the exaggerated lengthening of short vowels (tshba'al harakat), the lengthening of the long vowels (ziyadahal, madd), omission of short vowels (taqtl'al buro1), and the improper addition of short vowels (tahrikal harf al sakin) (al Sa'ld 1967:347)."

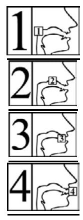
[9] The English translation reads: "Thus the LORD scattered them from there over the face of the whole earth; and they stopped building the city."

[10] "*Tajwīd* [is] the system of rules regulating the correct oral rendition of the Qur'an. The importance of *Tajwīd* to any study of the Qur'an cannot be overestimated: *Tajwīd* preserves the nature of a revelation whose meaning is expressed as much as by its sound as by its content and expression and guards it from distortion by a comprehensive set of regulations which govern many of the parameters of the sound production, such as duration of syllable, vocal timbre and pronunciation," [5, p. 16].

[11] "*Tartīl,* in the science of the Qur'ān, an incantatory mode of recitation." See "tartīl," in [16].



Example 8. Mouth and lips positions (in relation to consonants or vowels).

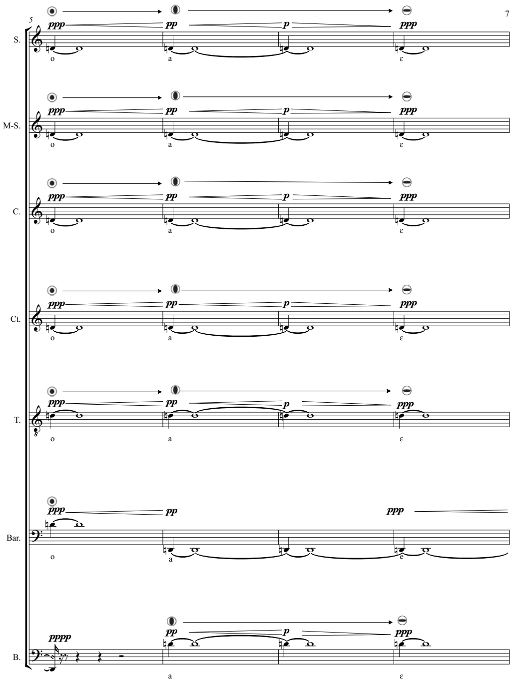


Example 9. Tongue positions in relation to mode of production and pronunciation

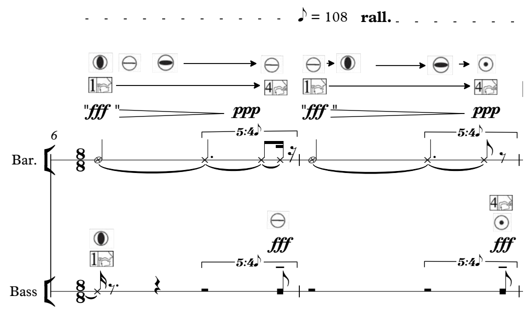
Qur'an, informed the creation of the notation, which indicates temporal elongation, constriction, acceleration and de-acceleration as well as mouth placement for correct Arabic pronunciation.

To emulate such techniques, a notational framework was created to allow the singers to correctly pronounce the Arabic text. Simultaneously, the notational framework provided the basis for new, yet "unheard" vocal sonorities. To this extent, two notational frameworks were employed for mouth and lip positions as well as tongue positions [17][12] (Examples 7 and 8).

While these notational frameworks were often used to emphasize pronunciation, they also allowed for totally "abstract" musical frameworks, far away from the world of sacred texts devoid of any concrete language. This was the case at the beginning of the composition *Al Ken Kara (The is Why It Was Called)*, where a kind of "pre-language" is created before the first enunciation of the Biblical text. Here, the notation of mouth positions was employed to ensure a gradual transition between vowel sounds and resulting changes in overtone spectra in many of the works.



Example 10. Notation of mouth positions for overtone singing at on the single tone of D3 at beginning of composition based on the text of the *Tower of Babel.*



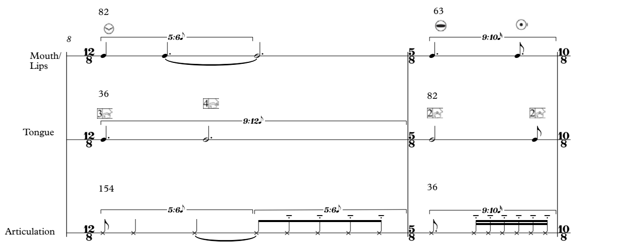
Example 11. Mouth positions combined with vocal techniques (whispering, inhaled whispering, throat tremolo) in mm. 6–7 in movement *Deus Sive Natura* in *Scholium Secundum* after Baruch Spinoza.

5. THE MUSIC OF NATURE: CONSTRUCTING SONOROUS STRUCTURE WITH GEMATRIA

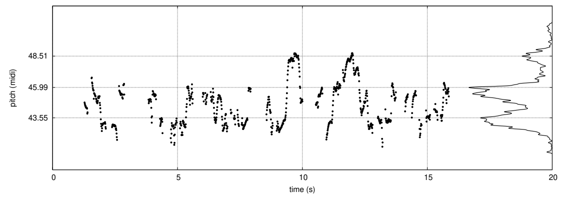
In the composition *Scholium Secundum* (2017–2022), notation was developed to create a sonorous framework that exists as an analogy to the concept of nature and the divine, as conceived by philosopher Baruch Spinoza (1632–1677) [10], [18, p. 296].[13] In the eighth movement of the composition, entitled *Deus Sive Natura,* a notational framework

[12] Such a notation was based on Aaron Cassidy's *A Painter of Figures in Rooms (2012).*

[13] In the fifth chapter of his *Ethics*, "On Human Freedom," Spinoza writes "For the eternal and infinite Being, which we call God or Nature, acts by the same necessity as that whereby it exists. For we have shown, that by the same necessity of its nature, whereby it exists, it likewise works (I. xvi.). The reason or cause why God or Nature exists, and the reason why he acts, are one and the same. Therefore, as he does not exist for the sake of an end, so neither does he act for the sake of an end; of his existence and of his action there is neither origin nor end."



Example 12. *Scholium Secundum* (2017–2022) Canon formed by combination of mouth positions tongue placement and articulation in movement *Deus Sive Natura,* after Baruch Spinoza.



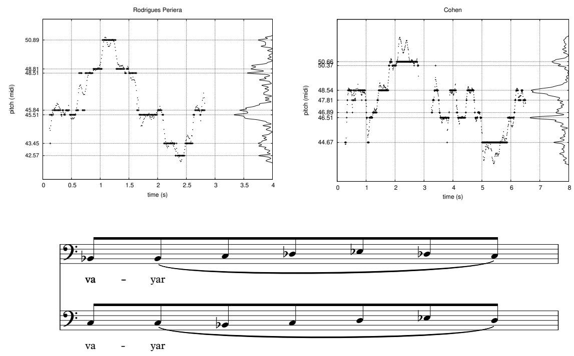
Example 13. Computational contour analysis of Torah recitation as practiced at the Portuguese Synagogue in Amsterdam.

was created to structure mouth and tongue positions, as various methods of sound production (whispering, inhaled, tremolo and vocal-fry techniques) were employed to produce an abstract "music of nature" following Spinoza's ideas.

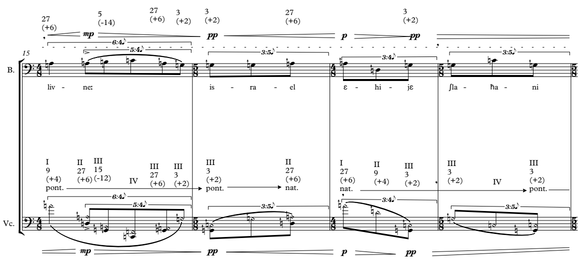
In this movement, the initial sketch was based on Gematria values of Spinoza's text [10], [19].[14] Here a canon, determined by the gematria values, initially forms the various durations of the three types of sound production (mouth/ lips, tongue and articulation).

6. HISTORICIZED COMPOSITION: MELODIC CONTOURS RE-MEMBERED

The pitch material for the twelve movements of this composition derives from melodies from the Portuguese synagogue in Amsterdam, the very synagogue from which Spinoza was excommunicated [20]. The computational analysis of the recorded recitation, as practiced at the Portuguese Synagogue in Amsterdam, presents not only the exact pitches but also the large-scale melodic contour. This melodic contour became integrated within an overtone-based harmonic framework of the composition.



Example 14. Analysis of melodic contour as practiced by two reciters associated with the Portuguese Synagogue in Amsterdam.



Example 15. Recomposed Torah melody in seventh movement of *Scholium Secundum* based on the computational contour analysis of Torah recitation as practiced at the Portuguese Synagogue in Amsterdam.

Recording various members of the Portuguese Synagogue in Amsterdam, we were able to compare differences between reciters. We were able to analyze how the Torah recitation melodies transformed over the course of the twentieth and twenty-first centuries. As illustrated in Example 14, a recitation practice from one practitioner is compared to another. We can thus perceive that, while both reciters employ a similar modal scale for their recitation, which is the normative framework for Torah recitation, the melodic gestures vary.

This contour analysis of this Torah melody formed the basis of the sixth and seventh movements of the composition. Here the tenor, which recites the melody, is accompanied by the cello, its varying string placement indications (tasto, naturale, ponticello, molto ponticello) providing a timbral analogy to vowels of the Hebrew text (u = tasto, a = nat. and i = pont) [10].

[14] Gematria was employed in terms of creating musical structures from both the Hebrew and Latin texts (using *Latin Qabalah Simplex).* The Latin gematria is based on premises of Hebrew Gematria: "The substitution of numbers for letters of the Hebrew alphabet, a favorite method of exegesis used by medieval Cabbalists to derive mystical insights into sacred writings or obtain new interpretations of the texts. Some condemned its use as mere toying with numbers, but others considered it a useful tool, especially when difficult or ambiguous texts otherwise failed to yield satisfactory analysis. Genesis 28:12, for example, relates that in a dream Jacob saw a ladder (Hebrew *sullam)* stretching from earth to heaven. Since the numerical value of the word *sullam* is 130 (60 + 30 + 40) – the same numerical value of Sinai (60 + 10 + 50 + 10) – exegetes concluded that the Law revealed to Moses on Mount Sinai is man's means of reaching heaven. Of the 22 letters in the Hebrew alphabet, the first ten are given number values consecutively from one to ten, the next eight from 20 to 90 in intervals of ten, while the final four letters equal 100, 200, 300, and 400, respectively. More complicated methods have been used, such as employing the squares of numbers or making a letter equivalent to its basic value plus all numbers preceding it."

7. CONCLUSIONS

In our analyses of Jewish, Islamic and Christian chant repertoires, we have set out to discover how salient pitch structures, contours and scale structures are determined and how these relate to the performance practice and larger cultural frameworks of these traditions. Simultaneously, the resulting computational analysis presents transcriptions of the chant repertoires. Combining computational analysis of resultant pitch structures determined by specific rules of Qur'an recitation and Torah Trope, we better understand how the rules for correct enunciation help to form and interact with given tonal hierarchies found within performances of both traditions. Simultaneously, we were able to create compositional sketches based on this material, allowing for musical structures wherein pronunciation, vocal timbre and sonority become main foreground parameters, allowing for new compositional strategies and frameworks for musical material development. Recent advances in deep learning and audio processing enable a less labor-intensive procedure to analyze and compare different chant recordings. In the next stage, we will employ automatic alignment of text and audio to bypass the need for manual annotation of the word and syllable boundaries in the audio files. Thus, we will be able to make comparisons on a larger scale. This will allow for a better understanding of these chant traditions, also allowing for new perspectives for recitation transcription and vocal composition.

Acknowledgments

We are grateful to the various reciters from Muslim and Jewish communities who have participated in this research as well as colleagues at our respective research institutions.

8. REFERENCES

[1] D. P. Biró, *Reading The Song: On the Development of Musical Syntax, Notation and Compositional Autonomy: A Comparative Study of Hungarian Siratók, Hebrew Bible Cantillation and Ninth Century Plainchant from St. Gallen.* Princeton University, 2004.

[2] P. van Kranenburg. *A Computational Approach to Content-Based Retrieval of Folk Song Melodies.* Ph.D. thesis. Utrecht: Utrecht University, 2010.

[3] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002. https://doi.org/10.1121/1.1458024

[4] H. Zimmermann, *Tora und Shira: Untersuchungen zur Musikauffassung des rabbinischen Judentums.* Bern: Peter Lang, 2000.

[5] K. Nelson, *The Art of Reciting the Koran.* Austin: University of Texas Press, 1985.

[6] H. Hucke, "Toward a New Historical View of Gregorian Chant," *Journal of the American Musicological Society, Volume 33,* Chicago: University of Chicago Press, 1980.

[7] D. P. Biró and P. van Kranenburg. "A Computational Re-Examination of Béla Bartók's Transcription Methods as Exemplified by his Sirató Transcriptions of 1937/1938 and their Relevance for Contemporary Methods of Computational Transcription of Qur'an Recitation," in *Proceedings of the Fourth International Workshop on Folk Music Analysis,* Istanbul, 2014.

[8] P. van Kranenburg, Dániel Péter Biró, Steven R. Ness, and George Tzanetakis, "A Computational Investigation of Melodic Contour Stability in Jewish Torah Trope Performance Traditions," in *Proceedings of the International Society on Music Information Retrieval (ISMIR2011) Conference*, pp. 163–168.

[9] D. P. Biró, *Al Ken Kara (That Is Why It Was Called).* Brühl, Edition Gravis, 2024.

[10] D. P. Biró, *Scholium Secundum.* Brühl, Edition Gravis, 2022.

[11] L. Ibsen al Faruqi, "The Cantillation of the Qur'an", *Asian Music*, Vol. 19, No. 1, 1987.

[12] A. Yūsuf Ali, *The Holy Qur'ān, Text, Translation and Commentary*, Amana Corp., 1989.

[13] B. Wheeler, "Arab Prophets of the Qur'an and Bible," *Journal of Qur'anic Studies*, vol. 8, no. 2, pp. 24–57.

[14] Rouen, Bibliothèque municipale, 248 (olim A. 339). Praha (Prague), Národní knihovna České republiky (National Library), XIV.B.13.

[15] R. Weber, *Biblia Sacra. Iuxta Vulgatam Versionem.* Stuttgart: Deutsche Bibelgesellschaft, 1994.

[16] P. J. Bearman et al. (eds.), *Encyclopaedia of Islam*, Second Edition, Brill Online, 2014, https://referenceworks.brill.com/display/db/ei1o (accessed April 2, 2024).

[17] A. Cassidy, "Constraint Schemata, Multi-axis Movement Modeling, and Unified, Multi-parametric Notation for Strings and Voices," in *Search, Journal for New Music and Culture* 10 (2013). Available: http://www.searchnewmusic.org/cassidy.pdf (accessed November 29, 2022)

[18] B. de Spinoza, *Ethics*. Introduction by S. Hampshire, transl. by Edwin Curley, London: Penguin Books, 1996.

[19] "Gematria" *Encyclopedia Britannica*, https://www.britannica.com/topic/gematria.

[20] A. Kasher and S. Biderman, "Why Was Baruch De Spinoza Excommunicated?", https://www.tau.ac.il/~kasher/pspin.htm (accessed on November 29, 2023).

GRANULAR POEM PROJECT: A DIGITAL SCORE BASED ON DECONSTRUCTION OF CHINESE IDIOMS

Zhang Yuan
zhang_yuan@ccom.edu.cn

Zhang Xinran
zxinran@ccom.edu.cn

Li Xiaobing
lxiaobing@ccom.edu.cn

ABSTRACT

We have established a reusable and iterative digital score [1] system, with Chinese four-character idioms at its core, enabling a collaboration between humans and artificial intelligence around selected idioms. Each entity engages in imagination and computation in its own unique way, facilitating spontaneous improvisation based on the mutual outputs. In this system, we have incorporated two models that we independently developed and trained: an AI-assisted modern Chinese poetry generation system and an AI-assisted piano MIDI generation system. Simultaneously, human participation in this musicking [2] process takes two forms. First, human performers interpret graphic scores designed for the selected idioms, engaging in improvisations. Second, audience representative contribute by providing real-time feedback through EEG brainwave signals, actively participating in the creative process. This system explores and expands the ways in which composers create with the assistance of computational tools. Humans and artificial intelligence, converging around a shared theme, dance in mutual reflection on the aesthetic level, making the process itself profoundly poetic. Also, it is the deep involvement of both humans and artificial intelligence that establishes this digital score and gives it its unique characteristics.

In this context, "deconstruction" refers to the process of subjecting the Chinese idioms selected in the Granular Poem Project to a series of direct and indirect processing steps, breaking them down into their constituent parts for the purpose of recombination, transformation, or reinterpretation. This process includes converting the idioms into modern poetry, musical fragments, and symbolically notated scores, as well as using these scores in improvised performances and selecting the next idiom to be processed based on brainwave activity. The aim of this deconstruction process is to reconsider and reinterpret the meaning and significance of idioms from different perspectives, in order to create new artistic forms and experiences.

1. INTRODUCTION

The Granular Poem Project aims to utilize a combination of technology and art, employing digital score as a medium to interpret and propagate Chinese culture. At its core, the project focuses on Chinese idioms, which constitute a unique form of fixed expressions typically comprised of four characters, each carrying distinctive meanings. Originating from ancient literary works, historical anecdotes, and folklore, idioms encompass profound cultural connotations, serving as integral components of the Chinese language and culture. In the Granular Poem Project, selected idioms, chosen as themes, provide essential material for deconstruction. The term 'granular' in the project title encapsulates the essence of its deconstructive processes.

Within the Granular Poem Project, chosen idioms undergo three direct deconstructions and two indirect deconstructions. The three direct deconstructions involve artificial intelligence generating modern poetry based on idioms, generating piano MIDI segments, and a composer disassembling and reassembling the Chinese characters' strokes of idioms into a graphic score. The two indirect deconstructions involve performers improvising based on the symbolic musical score, and audience wearing EEG devices selecting the next idiom for deconstruction based on their brainwave states during music appreciation. The multidimensional involvement of both artificial intelligence and human participants is a unique aspect of the Granular Poem Project as a musicking process.

The term "granular", borrowed from the technique of granular synthesis and incorporated into the project's title, succinctly encapsulates the deconstructive processes described above. In the realm of electronic music, granular synthesis involves generating small particles from a whole (deconstructing), which are then creatively reassembled by musicians to produce innovative sounds (poetic outcomes) [3]. In other words, the poetic essence of the term 'granular' lies in the freedom of imagination it grants to participants in this musicking process.

2. PROCESS OF DECONSTRUCTION

Firstly, idioms are concretely deconstructed in the process of generating poetry through artificial intelligence. Although idioms, consisting of four characters, are rich in meaning, their highly condensed form makes them abstract

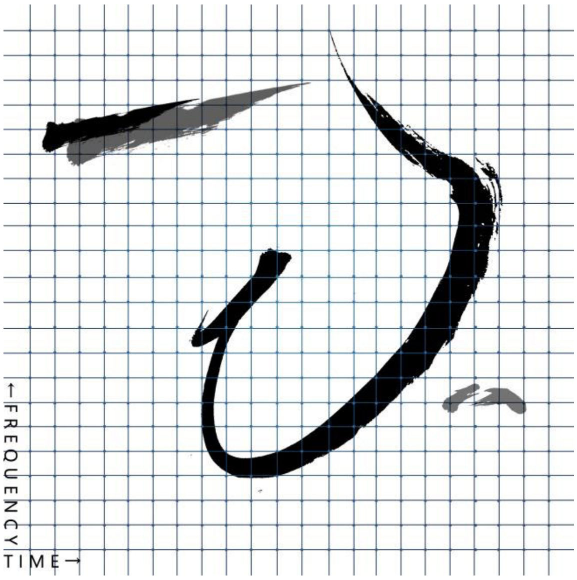


Figure 1. Example of graphical score made of calligraphy strokes. Each horizontal grid represents 2 seconds, and each vertical grid represents a range from 200 Hz to 1000 Hz, subject to the performer's discretion and the characteristics of the instrument.

and not easily understood. With the assistance of AI, idioms are treated as seeds that grow into modern poetry, complete with concrete scenes, images, and details. This transformation from abstract concepts to tangible elements imparts poetic qualities to idioms while making them more accessible.

The second form of deconstruction applied to the selected idioms is abstraction. Idioms are transformed into MIDI events and further extended through artificial intelligence to generate audio, becoming part of a musical performance. At this point, written symbols are converted into audible sounds, marking the second level of poeticization.

The third level of deconstruction is carried out through graphical symbolization. The brushstrokes of Chinese characters in calligraphy fonts inherently possess a dynamic quality in space and time. In this project, after deconstruction, transformation, and reassembly, these strokes are placed on a canvas with frequency and time domain coordinates. Performers can then engage in improvisational play based on their understanding of music and symbols.

In the example in Fig. 1, the intensity of stroke color represents the magnitude of amplitude. The vertical position of strokes in the composition signifies frequency, while the horizontal extension of strokes reflects the duration of time. These prescriptive representations provide the performer with a space for exploration of imagination, serving as a bridge to achieve the ultimate expression of sound.

Following these three direct deconstructions, there are two indirect deconstructions by performers and audiences: firstly, performers engage in improvisational play by reading graphic score derived from Chinese character strokes.

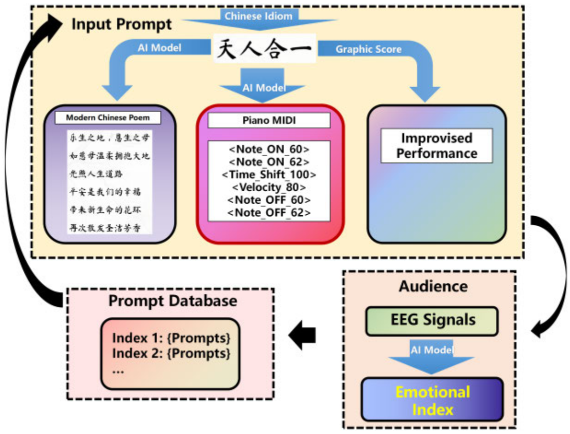


Figure 2. Digital Score framework of this project. The idiom we used is "天人合一" (oneness of heaven and man) [4]. The translation of the modern Chinese poem generated with AI assistance is as follows:

The land of joyous vitality,
The mother of grace and favor,
Like a gentle mother embracing the earth,
Lighting the path of human existence.
Peace is our happiness,
Bringing a garland of new life,
Once again radiating a fragrance of purity.

The sound produced during this improvisation, along with the sounds from the three direct deconstructions, is appreciated by the audience wearing EEG device. The audience's EEG signals, generated while appreciating the sounds, represent another indirect deconstruction of the idioms. The audience's EEG signals determine the next idiom to be deconstructed. Thus, the audience actively participates in the entire construction process of the music.

3. CONSTRUCTING THE SYSTEM

3.1. AI-Assisted Modern Poetry Generation

We adopt Lingxi, an AI-assisted modern Chinese poetry generation system [5] to generated poetry from the input prompt (Fig. 2). The system has the following features. For one hand, it incorporates an auto-regressive Transformer language model that was pretrained on a corpus containing more than 3,000 books of Chinese novels, and was fine-tuned on more than 220,000 passages of modern Chinese poetry. For another, it features in a novel Chinese word segmentation method as well as a set of novel sampling algorithms to promote the performance of the system. It also has a publicly available link (http://lingxi.website).

These above features provide convenience for our composition. We utilize the tokenization function of Lingxi to convert the input prompt into a series of tokens. These tokens are used as starting tokens for the language model to predict and generate remaining tokens, which is converted

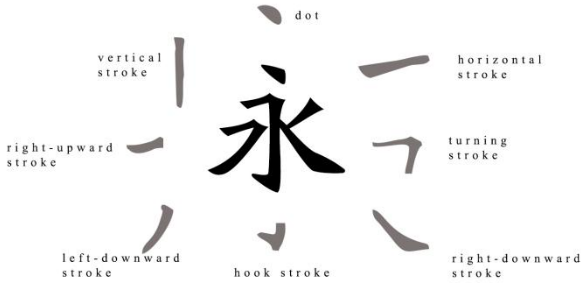


Figure 3. The character “永(yong)” demonstrates the eight fundamental strokes. The numbers in the picture above show the writing order of the strokes.

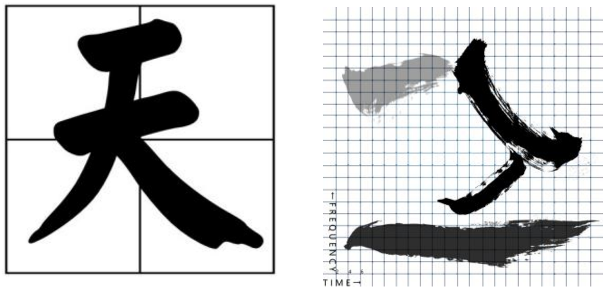


Figure 4. The printed form of the Chinese character “天 (tian)” (left) and the deconstructed brushstrokes as score (right).

back to a passage of Chinese poetry as the first type of interpretation of the prompt by the AI model.

3.2. AI-Assisted Piano MIDI Generation

Imitating the modern poetry generation paradigm, we train a parallel auto-regressive Transformer language model on the MAESTRO dataset [6]. It is a large scale piano MIDI music corpus with performance information. With the trained language model, it can generate piano MIDI music with performance information, upon which we utilize FluidSynth (https://www.fluidsynth.org/) to convert MIDI signals into audio files. To form a unified framework with poetry generation, we use the paradigm of natural language to build the tokenizer for piano MIDI music. We treat each MIDI event on the MAESTRO dataset as a single Chinese word, and use SentencePiece model [7] to aggregate high frequency combination of “words” into compound words. This results in a vocabulary of 32,000 size. With such vocabulary, the input prompt of the poetry can be mapped into MIDI event(s), upon which a MIDI signal can be generated by the trained language model. The final audio MIDI music file is regarded as the second type of interpretation of the prompt by the AI model.

3.3. AI-Assisted Emotion Detection From EEG Signal

The audience listens to the joint performance on the stage with an EEG equipment to record and send the EEG signal to the server. In such scenario, we seek to unearth the emotional information of the audience in the EEG signal. To detect the emotional index from the EEG, we train a classification model with a publicly available EEG dataset with emotion labels. With the trained model, we may infer the emotion of the audience from its EEG signal. Then, we use the inferred emotional index to determine a corresponding prompt from the candidates set, which serves as the input of the next round of performance.

3.4. Graphic Score Using Chinese Brushstrokes

The basic strokes of traditional Chinese characters are eight in number, namely dot (丶), horizontal stroke (一), turning stroke (㇕), vertical stroke (丨), hook stroke (亅), right-upward stroke (㇀), left-downward stroke (丿) and right-downward stroke (㇏). They can be demonstrated using the character "永" (yong) in the so-called formal script [8] (Fig. 3) .

Building upon these eight basic strokes, we aim to manifest the musicality and rhythmic essence inherent in Chinese character strokes by further transforming them and visually representing them in a graphic score.

For example, the horizontal stroke (一) exhibits natural variations in calligraphic writing, and if these variations are mapped onto sounds, it can evoke diverse associations and imaginations. Taking the character “天 (tian)” (sky/heaven) as an example, this Chinese character consists of two horizontal strokes. The inherent length difference between these two horizontal strokes can convey different musical implications. Building upon this, through processing in software, we can make the musical trends more explicitly enhanced. Additionally, there are two more strokes, namely "撇(pie) " left-downward stroke (丿) and "捺(na) " right-downward stroke (㇏). These two strokes exhibit a spatial relationship characterized by mutual support and opposing shapes. As these strokes undergo transformation and rearrangement on the canvas, they evolve into a prescriptive score with distinctive Chinese cultural characteristics (Fig. 4).

4. PERFORMANCE AND ITERATION

4.1. Performance Arrangement

In the Granular Poem Project, there exists an interconnected relationship among idioms, performers, and the audience, facilitated by the assistance of artificial intelligence, forming a chain of influence in the process of musicking.

Concerning idioms, we prepare a minimum of 12 idioms as a dataset for each performance. Each performance will commence with the idiom "天人合一" (oneness of heaven and man). "天人合一" is a philosophical concept rooted in traditional Chinese culture, integrating Daoist, Confucian, and Buddhist thoughts. It emphasizes the harmonious unity between humans and nature, advocating for the integration of humanity with the natural and cosmic realms.

Starting with this idiom represents the worldview of the creator.

As for performers, depending on the practical conditions of each performance, 1 to n performers can be invited to engage in live improvisation based on the graphic score. Simultaneously, the number of audience members wearing EEG collection devices can range from 1 to n.

4.2. Iteration of the Project

In each performance, the sequence of idiom appearances, the number of performers along with their individual interpretations, and the quantity of audience members, each with their own understanding and interpretations, all vary based on the circumstances, providing a fertile ground for the iteration of this project.

Regarding the selection of idioms, the time allocated for the presentation and interpretation of each idiom in every performance is adjusted to 1-3 minutes based on the EEG signals from the audience. Completing the presentation of 4 idioms is considered one full performance. In other words, with a dataset of 12 idioms, if 4 idioms are chosen for each performance, there are 495 different possible combinations. Additionally, more idioms are being deconstructed and added to the dataset.

We will conduct and document multiple performances, observing how different performance combinations influence the iteration, forming a data collection as case studies for the DigiScore project.

5. CONCLUSION

The goal of the Granular Poem Project is to merge technology and art, interpreting and disseminating Chinese culture through the medium of digital score. At its core, the project revolves around Chinese idioms and undergoes a multilayered process of deconstruction, employing methods such as artificial intelligence, improvisational performances, and graphic score to present a unique and creative artistic experience. Throughout the process of deconstruction, idioms transition from abstract linguistic forms to concrete poetic lines, then to audible audio expressions, ultimately manifesting on graphic score. This multidimensional deconstruction not only enriches the expressive forms of idioms but also provides audiences with diverse opportunities for participation, perception, and understanding.

The uniqueness of the Granular Poem Project also lies in its fusion of calligraphy and graphic symbols, along with the interaction between performers and audiences, creating a distinctive experience of musicking. Audiences are no longer mere passive listeners but actively engage in the music creation process through brainwave signals, organically contributing feedback for the selection of the next idiom.

The project incorporates fundamental Chinese brushstrokes and elements of Chinese culture, such as the introduction of the concept of "天人合一" (oneness of heaven and man), deeply intertwining with traditional Chinese philosophical thoughts. This cultural amalgamation positions the Granular Poem Project as not only an innovative blend of technology and art but also a modern interpretation of traditional culture.

The iterative nature of the Granular Poem Project turns each performance into a new exploration. The sequence of idiom appearances, the number of performers, and the level of audience participation all provide fertile ground for project innovation. The observations and recorded data from these performances will serve as valuable material for future research, exploring the impact and evolution of different performance combinations.

In summary, the Granular Poem Project is not just a creative practice in digital score, it is also a unique and profound exploration of Chinese idioms and their cultural connotations. Through this project, we witness the fascinating intersection of technology and art, as well as the organic fusion of Chinese and Western cultures in contemporary art. The reflections on different dimensions formed based on idioms are documented through sound, turning music into an embodiment of intellectual convergence, while also revealing the inherent poetry within [9].

Acknowledgments

We would like to express our sincere gratitude to the DigiScore project (ERC-2020-COG – 101002086) for their generous support of this research. We would also like to extend our heartfelt appreciation to Prof. Kenneth Fields for his ongoing encouragement and insightful feedback.

6. REFERENCES

[1] C. Vear, *The Digital Score: Musicianship, Creativity and Innovation.* New York: Routledge, 2019.

[2] C. Small, *Musicking: The meanings of performing and listening.* Wesleyan University Press, 1998.

[3] C. Roads, "Introduction to granular synthesis," *Computer Music Journal,* vol. 12, no. 2, pp. 11–13, 1988.

[4] L. Jiao, *A Cultural Dictionary of The Chinese Language: 500 Proverbs, Idioms and Maxims 文化五百条*. London: Routledge 2019. https://doi.org/10.4324/9780429356476.

[5] X. Zhang, M. Sun, J. Liu, X. Li, "Lingxi: A Diversity-aware Chinese Modern Poetry Generation System," in *Proc. ACL* (demo) 2023, pp. 63–75. https://doi.org/10.48550/arXiv.2108.12108.

[6] C. Hawthorne et al., "Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset," in *Proc. ICLR* 2019. https://doi.org/10.48550/arXiv.1810.12247.

[7] T. Kudo, J. Richardson, "SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing," in *Proc. EMNLP* (Demonstration) 2018, pp. 66–71.

[8] T. Chen, *Chinese Calligraphy.* Cambridge: Cambridge University Press, 2011.

[9] J. Liu, *The Art of Chinese Poetry.* London: Routledge, 2022.