# BEYOND THE BASIC-SPACE OF TONAL PITCH SPACE: DISTANCE IN CHORDS AND THEIR INTERPRETATION

**Hiroyuki Yamamoto**
JAIST
Ishikawa, Japan
yamamoto@kusuli.com

**Satoshi Tojo**
JAIST
Ishikawa, Japan
tojo@jaist.ac.jp

## ABSTRACT

Tonal Pitch Space (TPS) defines a numerical distance between two chord interpretations. Although it is based on musical knowledge and theory, the structure and values are not defined in an objective manner. Preceding works have addressed this problem, and TPS has been revised and optimized the definitions of distance, in the interpretation of chord paths, given chord names. But, because of the property of the task they used, they failed to reassess one of the three subelements of TPS, basic-space. In this study, we modify the task to incorporate pitch class (PC) information so that we can not only train other distance models that concern PC but also compare their performance with that of basic-space. We show that the data-oriented approach improves the accuracy from the original basic-space, especially when we add a distinction of major and minor keys.

## 1. INTRODUCTION

Tonality identification is an attractive but hard issue; although human listeners, often without any difficulty, fix one key to understand/ recognize tonal music, the exact process is still unknown. To determine a key, we need to consider the relationship between chords, considering cadences or tension/ relaxation structure. Moreover, we also need to model the relationship between each chord and pitch class (PC). But, when we represent this cognitive process in computers, we are required to assess the relationship objectively, excluding our subjectivity, so that the numeric distance in chords should be an intrinsic clue; *Tonal Pitch Space* (TPS) [8] has been one of the most convincing theories to give such a numeric distance between two chords.

Thus far, we have employed TPS to measure the distance in chords, however, some definitions of TPS look arbitrary. For example, the notion of *basic-space* (Figure 1) gives different hierarchical importance among 12 tones in an octave, diatonic (scale) tones, the third, the fifth, and the root in this order; but, is the difference of importance always one?

Yamamoto And Aojo [15] dared to Avoid employing the numerical definition of APS, Aut Anstead, they tried to make machines learn the similar distance model, giving a sequence of Berklee chord names. Since a chord name already includes rich information, it can be limitedly interpreted into pairs of key and degree without needing to consider the relationships to each PC. Therefore, they cannot reduce the definition of distance to each PC; that is, the adequacy of the basic-space in Figure 1 was untouched.

We reconsider the importance of the basic-space so that we once abandon chord names, excluding the bias offered by chord names, and employ chroma vectors, which directly mention each PC. With this, we try to obtain a statistic model which behaves similarly to TPS, to give a plausible interpretation for a sequence of pitch events.



## 2. PRELIMINARIES

### 2.1 Tonal Pitch Space

TPS is the quantitative harmony analysis proposed by Fred Lerdahl [8]. It is proposed to complement Lerdahl's original music theory, *A Generative Theory of Tonal Music* (GTTM) [7] which applies generative grammar to extend the Schenkerian theory. In TPS, a chord (e.g., C major triad) is interpreted as a pair of a key and a degree (e.g., interpretations of C major triad are as follows: I/C, III/a, V/F, IV/G, VI/e, and VII/d), then distances are defined between these chord interpretations. The distance between chord interpretations $x$ and $y$ can be calculated as (1).

$$\begin{aligned}\delta(x,y) =& \text{region}(x,y) + \text{chord}(x,y) \\ &+ \text{basic-space}(x,y)\end{aligned} \tag{1}$$

where $\text{region}(x,y)$ is a distance between keys, $\text{chord}(x,y)$ is a distance between degrees.

$\text{basic-space}(x,y)$ is a distance on a structure called basic-space which concerns the importance of each PC relating to the chord interpretations. Basic-space is composed of five levels (i.e., root, fifth, triad, diatonic, and octave) and each level contains the PCs reflecting the chord interpretations. Figure 1 shows the example when $x = \text{I/C}$ and $y = \text{iv/d}$. For each chord interpretation, the root PC of the chord has four circles (i.e., up to the root level), the fifth PC has three circles, the third PC has two circles, and every other diatonic PC has one circle [1]. Then the distance between two chord interpretations is defined as the number

[1] We omit the octave level in Figure 1 and Figure 3 because it does not affect the results.

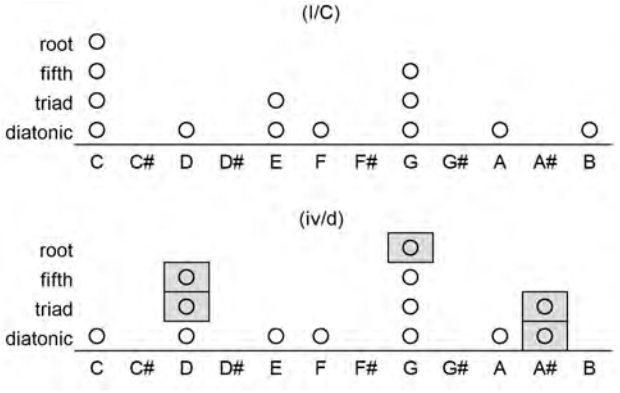


**Figure 1**. Basic-space.

of circles that exist only in the destination (the boxed circles in Figure 1). In this case, $\text{basic-space}(\text{I}/\text{C}, \text{iv}/\text{d}) = 5$. The details are explained in [8].

The calculation above is applicable only when $x$ and $y$ are in relative keys which are defined as follows:

$$\text{C}(R) = \begin{cases} \{\text{I}, \text{i}, \text{ii}, \text{iii}, \text{IV}, \text{V}, \text{vi}\} \\ \qquad \text{if } R \text{ is a major key} \\ \{\text{i}, \text{I}, \text{bIII}, \text{iv}, \text{v}, \text{bVI}, \text{bVII}\} \\ \qquad \text{otherwise} \end{cases} \tag{2}$$

where $\text{C}(R)$ is the set of all relative keys of key $R$.

If $x$ and $y$ are not in relative keys, distance between $x$ and $y$ can be calculated as:

$$\delta(x, y) = \min \begin{pmatrix} \delta(x, T_{R_1}) + \Delta(R_1, R_n) + \delta(T_{R_n}, y) \\ |R_1 \in \text{C}(R_x), R_n \in \text{C}(R_y) \end{pmatrix}$$

$$\Delta(R_1, R_n) = \min \begin{pmatrix} \sum_{i=1}^{n-1} \delta(T_{R_i}, T_{R_{i+1}}) \\ |R_{i+1} \in \text{C}(R_i) \end{pmatrix} \tag{3}$$

where $T_R$ is key $R$'s tonic, $R_z$ is chord interpretation $z$'s key. In other words, the transition from $x$ to $y$ must be considered as a combination of transitions within relative keys, and calculate the tonal distance for each combination, and then the shortest of these total distances is taken as the distance between $x$ and $y$.

### 2.2 Distance Models concerning Harmonic Features

There have been a lot of approaches to applying some kinds of space to model harmonic features and utilizing the distance to calculate plausibility. Heinichen [5], Kellner [6], and Weber [14] tried to define the space to express the positional relationships of each key area (region). Riemann [10] applied the Tonnetz, which had been invented by Euler [3] as a way of representing just intonation, to analyze harmonic relationships from the viewpoint of PC. Bharucha and Krumhansl [1] proposed a model of tonal hierarchy which has an empirically defined value to express relationships between chords within the same region. Randall et al. [9] explored the similarities with Lerdahl's TPS [8], which is defined rather theoretically, as a metric space, and proposed another distance model. Tymoczko et al. [12] formalized the levels of abstraction when we try to interpret harmony. Yamamoto and Tojo [15] generalized the structure of TPS and applied machine learning to train several distance structures. Here, we try to extend their approach further to complement their study.

## 3. OUR APPROACH

In this study, we aim to obtain optimal distances between PCs and chords, through the task of finding the most plausible path in chord interpretation, from chroma vectors. First, we review the issue of chord interpretation (§3.1). Second, we introduce some distance models which can calculate the distance between a chroma vector and a chord interpretation (§3.2). Then, we explain the way how to embed the models of §3.2 into the method of §3.1 to enable the method to receive chroma vectors instead of chord names (§3.3).

### 3.1 From Chord Names to Chord Interpretation Paths

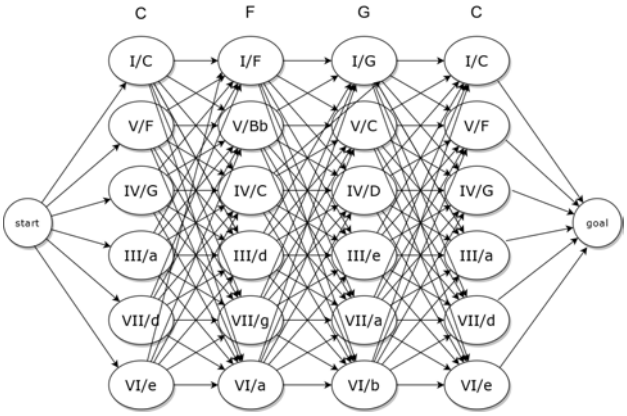


**Figure 2**. Interpretation graph.

Sakamoto et al. [11] have proposed a method to find the most plausible interpretation of a given chord name sequence. Given a chord name sequence, first, their method extends each chord to its interpretations and constructs a graph whose edges have weights that correspond to the distances on TPS. Then it applies the Viterbi algorithm to find the shortest interpretation paths from the start to the goal. Figure 2 shows an interpretation graph for chord name sequence $\text{C} \rightarrow \text{F} \rightarrow \text{G} \rightarrow \text{C}$. One of the shortest interpretation paths in Figure 2 is $\text{I}/\text{C} \rightarrow \text{IV}/\text{C} \rightarrow \text{V}/\text{C} \rightarrow \text{I}/\text{C}$.

Yamamoto and Tojo [15] have tried to generalize TPS and proposed several functions called "distance elements (DEs)" and a way to train them with annotated datasets. Based on the method of [11], their method replaces the TPS with the proposed generalized TPS then convert path distance to path probability such that the shortest path should have the highest probability, and finally apply SGD to update parameters. Their best model (i.e., a DE or combination of DEs) achieved over 86% accuracy while the original TPS was about 40%, and they also found a model with just 58 learnable parameters could achieve more than 80%.

In this study, we pick one of the most effective DEs proposed in [15] [2] as the base model on which we extend the structure in §3.3.

### 3.2 Between Chroma Vectors and Chord Interpretations

In this section, we introduce chroma distance models which are inspired by the structure of basic-space (and the basic-space itself is one of them).

A chroma vector is a 12-dimensional vector, the element of which represents its membership (1/0) or graded salience of the corresponding PC. We define the distance between a chroma vector and a chord interpretation as the sum of all distances between PCs and the chord interpretation.

Firstly, we can calculate this distance using basic-space. For example, the distance between the chroma vector [1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0] and the chord interpretation I/C can be calculated as the inner product of the chroma vector and the vector generated from the basic-space (as the number of gray boxes) as in Figure 3. This means basic-space divides PCs into five categories, namely, root (i.e., C in this case), third (i.e., E), fifth (i.e., G), diatonic (i.e., D, F, A, B), and the others then gives the predefined PC-level distance values as in Table 1.

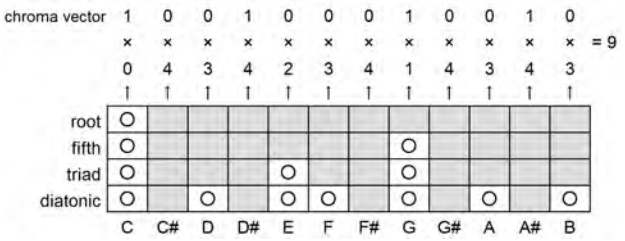


**Figure 3**. The distance between [1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0] and I/C based on basic-space.

| root | third | fifth | diatonic | other |
|---|---|---|---|---|
| 0 | 2 | 1 | 3 | 4 |

**Table 1**. The PC-level distance values from basic-space.

But, how to classify PCs and what distance values to apply are not obvious. So we try other possible models. Although the distance values are predefined in the original basic-space, the distance values in the following models will be learned by machine learning given an annotated dataset.

The first model, ch_dist_2, simply considers if the PC is in the chord note (i.e., root, third, and fifth) or not. The next model, ch_dist_3, distinguishes whether the PC is diatonic or not in addition to the distinction by chord membership. The next one, ch_dist_5, uses the same categories as those of basic-space. And finally, ch_dist_10 uses the same five categories but also distinguishes major or minor. We also define a dummy model, ch_dist_0, for comparison. This one always returns 0 regardless of what input is given. Table 2 shows the chroma distance models defined above.

[2] DE 8.1. This one achieved 86.25% accuracy with 686 parameters.

| | PC classification | params |
|---|---|---|
| **basic-space** | root/third/fifth/diatonic/other | 0 |
| **ch_dist_0** | - | 0 |
| **ch_dist_2** | chord/other | 2 |
| **ch_dist_3** | chord/diatonic/other | 3 |
| **ch_dist_5** | root/third/fifth/diatonic/other | 5 |
| **ch_dist_10** | (root/third/fifth/diatonic/other) ×(major/minor) | 10 |

**Table 2**. Chroma distance models. **params** is the number of learnable parameters.

### 3.3 From Chroma Vectors to Chord Interpretation Paths

The method explained in §3.1 receives chord names as the input, but now we modify it to receive chroma vectors. The graph structure becomes like Figure 4. The layer width becomes 24 (keys) × 7 (degrees) = 168 because all interpretations should be considered at every layer. Then every layer is duplicated to accept chroma vector inputs. Nodes in duplicated layers are connected by horizontal edges whose weights express the distances in the models introduced in §3.2.

The learnable parameters are trained to maximize the path probability of the ground truth paths. However, the formula of path probability is revised as follows because of the modifications in the interpretation graph.

$$\mathrm{P}(X_{0:s} = x_{0:s} | c_{0:s}, G_{0:2s}) \triangleq \begin{cases} 1 & \text{if } s = 0 \text{ [3]} \\ (\prod_{t=0}^{s-1} \frac{\exp(-\mathrm{CD}(c_t, x_t) + \mathrm{GTPS}(x_t, x_{t+1}))}{Z_{c,G,t}}) & \\ \times \mathrm{CD}(c_s, x_s) / Z^{(2)}_{c,G,s} & \text{otherwise} \end{cases} \quad (4)$$

where

$$Z_{c,G,t} \triangleq \sum_{l \in G_t} \sum_{m \in G_{t+1}} \mathrm{P}(X_t = l | G_{0:2t-1}) \times \exp(-\mathrm{CD}(c_t, l) + \mathrm{GTPS}(l, m)),$$

$$Z^{(2)}_{c,G,t} \triangleq \sum_{l \in G_t} \mathrm{P}(X_t = l | G_{0:2t-1}) \exp(-\mathrm{CD}(c_t, l)),$$

$x_t$ is a chord interpretation at $t$, $c_t$ is a chroma vector at $t$, CD is a chroma distance model, and GTPS is a generalized TPS proposed in [15]. This formula is designed to convert a path distance (i.e., $\sum_{t=0}^{s-1}(\mathrm{CD}(c_t, x_t) + \mathrm{GTPS}(x_t, x_{t+1}))$) to a path probability so that the shorter (shortest) path has higher (highest) probability.

The new graph (Figure 4) is a lot more complex than the original graph (Figure 2). But, all layers (except for the start and end layers) have the same set of nodes so all edges are the same too. Especially, even if a set of fully-connect edges become $168 \times 168 = 28,224$ (originally it was $6 \times 6 = 36$), it can be utilized repetitively. Moreover, the edges of chroma distances (i.e., the horizontal edges

[3] 0th layer contains only one node, that is, the start node

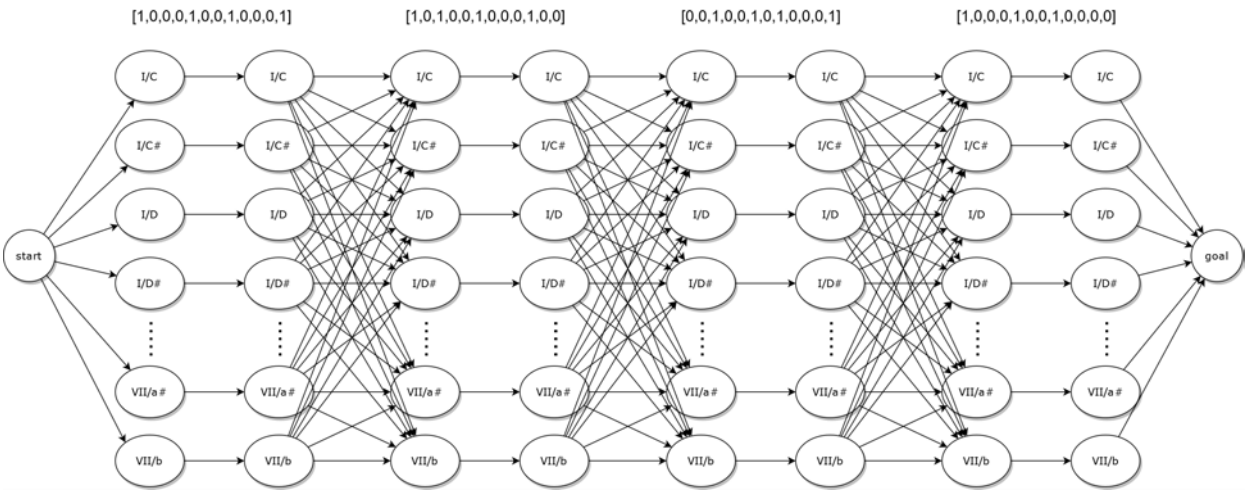


**Figure 4**. Revised interpretation graph.

below chroma vectors) can be calculated by a matrix product. Therefore, the computational cost does not increase as it looks.

## 4. EXPERIMENTS

### 4.1 Dataset

We use the dataset annotated in `rntxt` format [13], published at [4]. There are 384 pieces (1,905 phrases, 68,463 chords) and we regard every phrase as a unit (i.e., to which we predict the interpretation sequences) but when a phrase exceeds 50 chords we divide it into units each of which does not exceed 50 chords. Then we use 80% for training, 10% for validation, and the remaining 10% for the test.

We extracted key, degree, and applied chord information from `rntxt`, then omit all repetitions of the same chord interpretations. About applied chords, a tonic chord is added at the end of every local key section to express pivot chord modulation. Chroma vectors are obtained from `rntxt` using music21 library [2].

We set all the initial parameter values to be zero and train them by mini-batch stochastic gradient descent with batch size=100 and learning rate=0.001. We continue training at least 10 epochs and until no accuracy update in the validation set for an epoch [4] then pick the parameter which gives the highest validation accuracy.

### 4.2 Results

Table 3 shows the performance of, and Tables 4, 5, 6, and 7 shows the resulting PC-level distance values of the chroma distance models defined in §3.2, and Table 8 illustrates distance values between chroma vector [1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1] and some chord interpretations by the distance models. The performance is evaluated by how frequently the found path goes through the ground truth node (i.e., chord interpretation) in the revised interpretation graph. **acc** shows the accuracy the method could estimate ground truth chord interpretation for each chroma vector. **key acc** shows the accuracy the method could estimate at the ground truth key for each chroma vector.

| | acc | key acc |
|---|---|---|
| **basic-space** | 50.21% | 60.14% |
| **ch_dist_0** | 3.58% | 11.22% |
| **ch_dist_2** | 49.30% | 57.07% |
| **ch_dist_3** | 49.00% | 58.67% |
| **ch_dist_5** | 53.52% | 62.86% |
| **ch_dist_10** | 55.53% | 65.67% |

**Table 3**. Performance of chroma distance models.

| chord | other |
|---|---|
| 0 | 2.4576 |

**Table 4**. Resulting PC-level distance values of ch_dist_2.

| chord | scale | other |
|---|---|---|
| 0 | 2.0414 | 2.6578 |

**Table 5**. Resulting PC-level distance values of ch_dist_3.

| root | third | fifth | scale | other |
|---|---|---|---|---|
| 0.8525 | 2.8191 | 0 | 3.1986 | 4.2753 |

**Table 6**. Resulting PC-level distance values of ch_dist_5.

| | root | third | fifth | scale | other |
|---|---|---|---|---|---|
| **minor** | 1.2041 | 2.5036 | 0 | 3.3633 | 3.6971 |
| **major** | 0.7070 | 2.6557 | 0.0956 | 3.1862 | 6.2754 |

**Table 7**. Resulting PC-level distance values of ch_dist_10.

Even if we used a fairly strong model (i.e., DE 8.1 from [15]), it is almost impossible to narrow down the candidates without a hint from chroma vector (i.e., ch_dist_0) [5] .

[4] We loosened the stopping condition because the original condition in [15] was too costly to conduct an exhaustive evaluation.

[5] Having said that, $3.58\%$ is much better than $1/168 \approx 0.60\%$

| | I/C | i/e | vi/C | VII/d |
|---|---|---|---|---|
| **basic-space** | 6 | 6 | 9 | 7 |
| **ch_dist_0** | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| **ch_dist_2** (Table 4) | 2.4576 | 2.4576 | 4.9152 | 2.4576 |
| **ch_dist_3** (Table 5) | 2.0414 | 2.0414 | 4.0828 | 2.6578 |
| **ch_dist_5** (Table 6) | 6.8702 | 6.8702 | 9.2163 | 7.9469 |
| **ch_dist_10** (Table 7) | 7.0710 | 6.6445 | 18.2302 | 9.7337 |

**Table 8**. Distance values between chroma vector [1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1] and some chord interpretations.

Compared to ch_dist_0, ch_dist_2 performed very well with only distinguishing the membership of chords. Also, separating diatonic PCs (ch_dist_3), relative positions in triad (ch_dist_5), and major or minor key (ch_dist_10) all contributed to improve accuracy to some extent. Moreover, the result shows that basic-space worked quite well. It went below the same category model (i.e., ch_dist_5) but outperformed fewer category models (i.e., ch_dist_2 and ch_dist_3). This result we think indicates the importance of distinguishing that five categories. The most complex model (ch_dist_10) can be thought as a combination of two revised basic-spaces for major key and minor key respectively. And it achieved the best performance.

Looking at the learned PC-level distance values, "other" category has the largest and "scale" category has second-largest values. This is consistent with basic-space (Table 1). But within the "chord" category, "root" has the smallest value in basic-space while "fifth" has the smallest in the learned values. It is surprising that giving "fifth" smaller value than "root" enables the method to find better interpretation paths. We think this needs to be investigated further.

## 5. CONCLUSION

In this research, we have reconstructed the theory of distance between chords, motivated by Tonal Pitch Space (TPS), and proposed a model to guess interpretations (pairs of key and degree) to a sequence of chords, represented by chroma vectors. Since chroma vectors do not refer to human-recognizable interpretation but mention only pitch classes, we can objectively compare the relation between the role of the basic-space in a key and distances in notes.

We have compared six different sets of parameters, including the raw basic-space, and proved that these stochastic models outperformed the original TPS. We have experimented music pieces upon an open database, and the data-driven distance learning improved the accuracy by five percent or so, especially when we added a distinction of major and minor keys.

Our future work includes further refinement of this stochastic TPS, adding other features such as musical *genre* or difference of age, and so on.


### Acknowledgments

This research was supported by JSPS Kaken 20H04302 and 21H03572.


## 6. REFERENCES


[1] J. Bharucha and C. Krumhansl, "The representation of harmonic structures in music: Hierarchies of stability as a functions of context", *Cognition*, 13(1), pp.63-103, 1983.

[2] M. S. Cuthbert, C. Ariza, "music21: A toolkit for computer-aided musicology and symbolic music data", in *Proc. of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pp.637-642, 2010.

[3] L. Euler, *Tentamen novae theoriae musicae*. St. Petersburg Academy, 1739.

[4] M. Gotham, R. Kleinertz, C. Weiss, M. Müller, S. Klauk. "What if the 'When' implies the 'What'?: human harmonic analysis datasets clarify the relative role of the separate steps in automatic tonal analysis", in *Proc. of the 22nd International Society for Music Information Retrieval Conference (ISMIR)*, pp.229–236, 2021

[5] J. D. Heinichen, *General-bass in der composition*. Dresden: J. D. Heinichen, 1728.

[6] D. Kellner, *Treulicher Unterricht im General-Bass*. Hamburg: C. Herold, 1737

[7] F. Lerdahl, R. Jackendoff, *A Generative Theory of tonal music*. Cambridge, MA, 1983.

[8] F. Lerdahl, *Tonal pitch space*. New York, Oxford University Press, 2001.

[9] R. R. Randall, B. Khan, "Lerdahl's Tonal Pitch Space Model and Associated Metric Spaces", in *Journal of Mathematics and Music*, 4, pp.121-131, 2010.

[10] H. Riemann, *Grosse kompositionslehre, Vol. 1*. Berlin: W. Spemann, 1902.

[11] S. Sakamoto, S. Arn, M. Matsubara, S. Tojo, "Harmonic analysis based on tonal pitch space", in *Proc. of the 8th International Conference on Knowledge and Systems Engineering (KSE)*, pp.230-233, 2016.

[12] C. Callender, I. Quinn, D. Tymoczko, "Generalized Voice-Leading Spaces", in *Science*, 320, pp.346-348, 2008.

[13] D. Tymoczko, M. Gotham, M. S. Cuthbert, C. Ariza, "The romantext format: a flexible and standard method for representing Roman numeral analyses", in *Proc. of the 20th International Society for Music Information Retrieval Conference (ISMIR)*, pp.123-129, 2019.

[14] G. Weber, *Versuch einer geordneten theorie der tonsetzkunst*. Mainz: B. Schotts Söhne, 1821-24.

[15] H. Yamamoto, S. Tojo, "Generalized tonal pitch space with empirical training", in *Proc. of the 18th Sound and Music Computing Conference (SMC)*, 2021, pp.300-307.