

Distance in Pitch Sensitive Time-span Tree

Masaki Matsubara

University of Tsukuba

masaki@slis.tsukuba.ac.jp

Keiji Hirata

Future University Hakodate

hirata@fun.ac.jp

Satoshi Tojo

JAIST

tojo@jaist.ac.jp

ABSTRACT

The time-span tree of Jackendoff and Lerdahl's Generative Theory of Tonal Music is one of the most promising representations of human cognition of music. In order to show this, we compare the distance in trees and psychological dissimilarity by using variations of *Ah vous dirais-je, maman* by Mozart. Since pitch and chord sequence also affect the time spans, we amend the time-span analysis to include pitch information. Then, we introduce a pitch distance based on Lerdahl's theory and revise the tree distance. We compare analyses with and without the pitch information and show the efficacy of our method.

1. INTRODUCTION

Cognitive similarity is one of the most important aspects of music, both for practical applications such as music retrieval, classification, and recommendation [15, 5, 17], and for modeling the human cognitive process [2, 3]. There are various viewpoints on evaluating this similarity, including melodic segmentation/parallelism, phonetic chromatography, and so on. In this paper, we consider structural similarity. Schenkerian Theory in the 1920's [13] put forward the *reduction hypothesis*; that is, the importance of each pitch event is different in a piece of music, and hence, we can retrieve an intrinsic skeleton of the music by picking these important events.

Although the idea of reduction starts with Schenker, there have been various approaches to reduction, such as Gestalt, grammatical, and memory-based models [4, 1, 10]. Among them, the *time-span analysis* in Lerdahl and Jackendoff's Generative Theory of Tonal Music (GTTM; hereafter) [11] avoids metaphysical issues and gives instead a more concrete process of reduction that is based on rhythmic and harmonic stability. The theory assigns a structural importance to each pitch event, derived by *grouping analysis* and *metrical analysis*. As neighboring events can be compared by using this structural importance, a branch from a less important event is absorbed into that from a more important event; as a result, a hierarchical structure forms a *time-span tree* in a bottom-up way (Figure 1.).

In the GTTM analysis, as the preference rules are rather arbitrarily defined, contrary to the well-formedness rules,

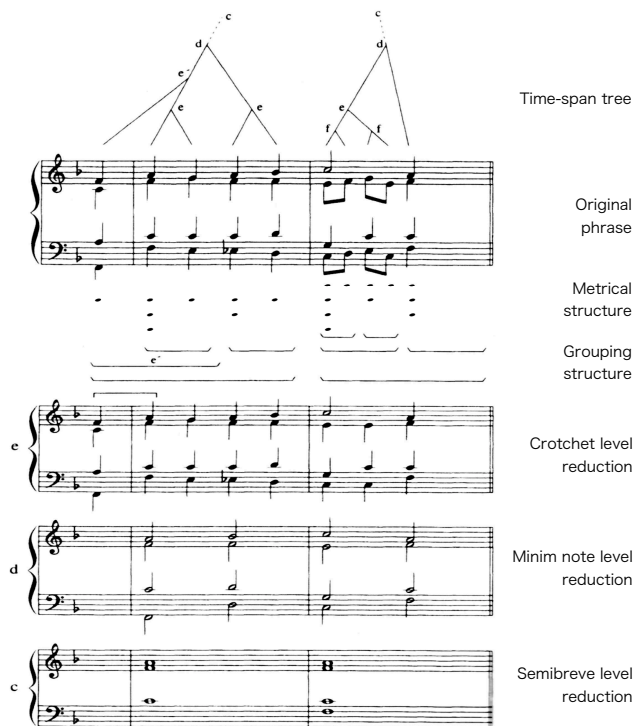


Figure 1. Time-span reduction of the first phrase of BWV281 [12, pp. 10–11].

they often conflict with each other. Hamanaka et al. [6] assigned parametric weights to each rule to control the process to avoid this problem, but the time-span tree still needs to be redressed by pitch and/or chordal information, which especially appear in half cadence or cadential retention.¹ In this paper, to amend the default of pitch information, we introduce a new preference rule based on Tonal Pitch Space (TPS; hereafter) [12].

Thus far, we defined the edit distance of a time-span tree [19] and measured the distance between variations of *Ah, vous dirais-je, maman* by Wolfgang Amadeus Mozart, K.265/300e [9], where the distance correctly reflects human intuition. One problem was that if one of the two variations was in a minor key, the rhythmic resemblance did not match the psychological similarity. In this paper, we tackle the same set of variations and show that the pitch information improves the situation.

This paper is organized as follows. In Section 2, we define the editing procedure of time-span tree together with

¹ The theory describes another tree, called the *prolongation tree*, which properly reflects the harmonic structure

the notion of maximal time span. In Section 3, we show our revision; we formally define the distance regarding the preorder of pitches and chords. In Section 4, we compare the results of our distance calculation with psychological similarity. In Section 5, we summarize our contribution and discuss future work.

2. DISTANCE IN TREE WITHOUT PITCH INFORMATION

We hypothesize that if a branch with a single pitch event is removed from a time-span tree, an amount of information proportional to the length of its time span is lost. The *head* pitch event of a tree is the most salient event of the whole tree; then, we may regard that its saliency is extended to the whole tree. The situation is the same as the head of each subtree. Thus, we consider that each pitch event has its maximal length of saliency, called the *maximal time span*.

Let $\zeta(\sigma)$ be a set of pitch events in σ and $mts(e)$ be the maximal time span of event e . For each reduction step, when event e on the reducible branch disappears, the length of its maximal time span $mts(e)$ becomes the distance of the step. The same goes for addition of a branch. Therefore, the distance d between two time-span trees, σ_A and σ_B , is defined by

$$d(\sigma_A, \sigma_B) = \sum_{e \in |\zeta(\sigma_A) - \zeta(\sigma_B)|} mts(e).^2$$

Note that there is a latent order in the addition and reduction of branches, though the distance is defined as a simple summation of maximal time spans. Finally, we can easily show the triangle inequality [19]:

$$d(\sigma_A, \sigma_B) + d(\sigma_B, \sigma_C) \geq d(\sigma_A, \sigma_C).$$

3. DISTANCE WITH PITCH INFORMATION

In the time-span reduction, there are several preference rules concerning pitch and harmony in GTTM. Of these, we will focus on TSRPR (Time-Span Reduction Preference Rule) 2 (Local Harmony).³ We assume that the relative consonance could be evaluated with the root note and chord inversion type. Thus, we redefine TSRPR2', as follows:

TSRPR2' (Local Harmony)

(a) prefer chord inversion as follows:

$$I > I^6 > I_4^6.$$

(b) prefer a chord that is relatively closely related to the local tonic as follows:

$$I > V > IV > VII > II > III > VI.$$

Dissonant notes⁴ often appear in a local harmony, and thus, we add a new preference rule based on TPS[12].

² $|A - B| \equiv A \cup B - A \cap B$.

³ "Of the possible choices for the head of a time-span T, prefer a choice that is (a) relatively intrinsically consonant, (b) relatively closely related to the local tonic."

⁴ as anticipation, neighbor tone, passing tone, etc.

TSRPR10 (New) (Local Pitch Consonance)

prefer pitch class in a local harmony as follows:

$$0 > 7 > 4 > \{2, 5, 9, 11\} > \{1, 3, 6, 8, 10\},$$

where each number represents the pitch class in the local key, e.g., if in G major the numbers are interpreted as $G > D > H$, and so on. Note that there is no preference among pitch classes in a brace.

Now, we define the pitch-sensitive distance. The distance is basically the edit distance weighted by the *maximal time span* introduced in Section 2. Some algebraic features of the distance are described in [19].

Tree Distance with pitch information

Let σ_A, σ_B be trees; the revised distance $d^\pi(\sigma_A, \sigma_B)$ is defined as follows.

$$d^\pi(\sigma_A, \sigma_B) = \sum_{e_j \in |\zeta(\sigma_A) - \zeta(\sigma_B)|} (\delta_{e_i}(e_j) \times mts(e_j)),$$

where $\delta_{e_i}(e_j)$ is the proximity from the pitch event on the parent branch e_i to that on the subordinate branch e_j .

We calculate the proximity based on TPS (Table 1)[12].

Let $d^\pi(\sigma_A, \sigma_B) = 0$ when σ_A and σ_B have only one pitch event each, but with different pitch classes of the same duration (shifting root).

For example, Figure 2 shows a calculation of the distance between melody C-F-A and melody C-G#-A. The distance is the difference of an F note which is to be removed from melody C-F-A (= 0.75), plus that of the G# note to be added to melody C-A (= 0.625), which results in a total of 1.375. Figure 3 also shows the tree distance of root shifting when no common note exists between the two trees.

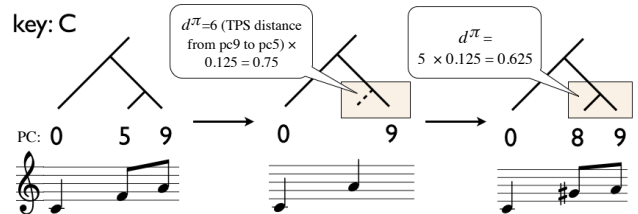


Figure 2. Pitch-sensitive tree distance (1.375 in total)

4. EXPERIMENTAL RESULTS

4.1 Materials and Methods

We experimented with different distances on the same material [9], that is, variations of *Ah, vous dirai-je, maman* by Wolfgang Amadeus Mozart K.265/300e (Figure 4). Although the original piece consists of two voices, we extracted a more salient pitch event between the two, as well as a prominent note per chord, and arranged the piece into a monophonic melody. In this process, we disregarded differences of an octave so that the resultant melody would be easier to hear.

Table 1. Pitch class proximity in TPS ([12, p. 49])

Pitch class (pc)	0	1	2	3	4	5	6	7	8	9	10	11
distance from pc0	0	5	4	6	3	5	7	2	6	5	6	4

Table 2. Tree Distance

	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10	No. 11	No. 12
Theme	13.31	33.0	18.42	34.92	8.88	32.94	13.44	19.25	11.25	47.06	26.5	51.63
No. 1	-	44.81	31.23	47.73	20.94	45.75	25.75	32.06	24.06	59.88	39.31	64.44
No. 2	-	-	44.92	18.92	41.38	37.44	43.94	43.75	39.75	51.56	42.88	56.13
No. 3	-	-	-	45.17	26.79	44.85	29.35	37.17	25.17	58.98	40.33	63.54
No. 4	-	-	-	-	43.29	28.69	45.85	45.67	41.67	53.48	44.71	58.04
No. 5	-	-	-	-	-	41.31	21.81	27.63	19.63	55.44	34.88	60.0
No. 6	-	-	-	-	-	-	43.88	43.69	39.69	51.5	42.81	56.06
No. 7	-	-	-	-	-	-	-	32.19	24.19	58.0	39.44	62.56
No. 8	-	-	-	-	-	-	-	-	27.5	57.81	41.25	62.38
No. 9	-	-	-	-	-	-	-	-	-	53.81	33.25	58.38
No. 10	-	-	-	-	-	-	-	-	-	-	56.94	70.19
No. 11	-	-	-	-	-	-	-	-	-	-	-	61.5

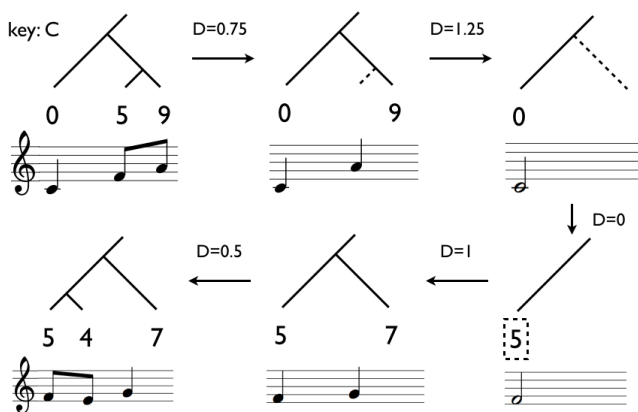


Figure 3. Distance including root shifting (3.5 in total)

First, we manually created the time-span trees of the theme and its twelve variations and cross-checked them. We made a chord sequence only on first eight-bars for each variation, with the help of a professional composer. The distance between two variations were calculated according to the definition in Section 3, including the new criteria of pitch difference. The number of comparisons amounted to 78 ($= {}_{13}C_2$) pairs.

Thereafter, we investigated the cognitive similarity; the examinees consisted of eleven university students, seven out of whom had experience in playing music instruments. The examinees listened to all the pairs $\langle m_i, m_j \rangle$ in random order without duplication, where $m_{\{i,j\}}$ was either the theme or variations No. 1 to 12. To cancel the cold start bias, the examinees listened to the whole theme and twelve variations (eight-bars long) without rating them. After that, each of them rated the intuitive similarity in five grades: $\{-2, -1, 0, 1, 2\}$. If one rated a pair of $\langle m_i, m_j \rangle$, he/she also tried the same pair later again in reverse order as $\langle m_j, m_i \rangle$ to avoid the order effect. Finally, the average ratings were normalized within all the examinees.

Figure 4. Monophonic melodies arranged for the experiment

4.2 Results

The experimental results are shown in the distance matrix in Table 2. Since the values of $d^\pi(\sigma_{m_i}, \sigma_{m_j})$ and $d^\pi(\sigma_{m_j}, \sigma_{m_i})$ are exactly the same, only the upper triangle is shown. The results of a conventional study, in which

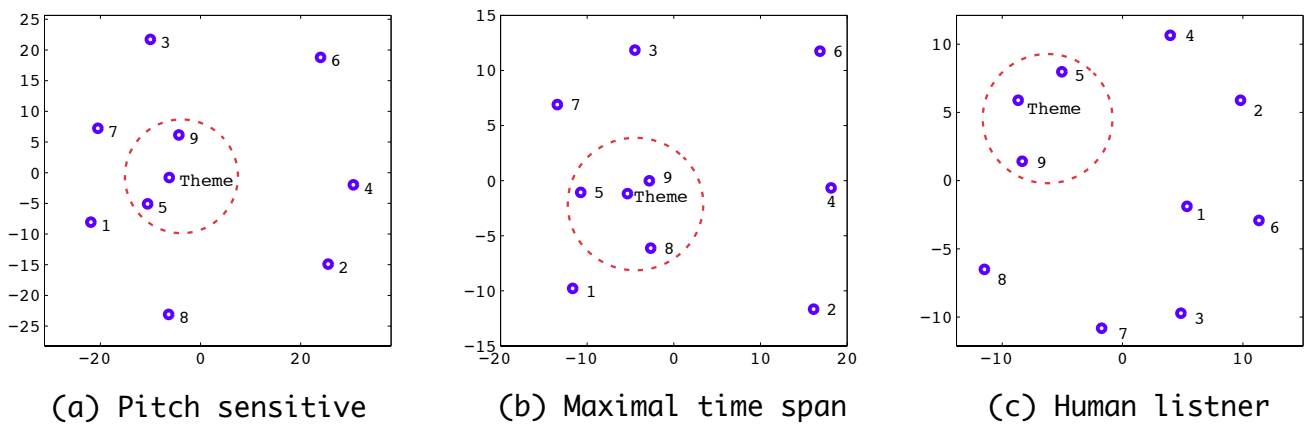


Figure 5. Relative distances among melodies in multidimensional scaling: (a) pitch sensitive and (b) only maximal time span (c) human listeners

examinees rated the psychological resemblance, are listed in Table 3 in the Appendix.

We employed multidimensional scaling (MDS) [20] to visualize the comparison. MDS takes a distance matrix containing dissimilarity values or distances among items, identifies the axes to discriminate items most prominently, and plots items on a coordinate system with the axes. In short, the more similar the items are, the closer they lie on the coordinate plane.

First, we used the MATLAB `mdscale` function, which uses Torgerson scaling of MDS, to plot the proximities of the 13 melodies; however, it was still difficult to find a clear distinction. Therefore, we restricted the target melodies to the theme and variations No. 1 to 9, as shown in Figure 5. The theme and No. i ($i = 1, \dots, 9$). in the figure correspond to those in Figure 5. The contributions in MDS were as follows: (a) tree distance with pitch information: first axis (horizontal) = 0.28, second = 0.20; (b) tree distance without pitch information: first axis (horizontal) = 0.23, second = 0.21; (c) human listeners: first axis (horizontal) = 0.33, second = 0.17.

4.3 Analysis

Here, we summarize the characteristic phenomena appearing in Figure 5.

Theme, No. 5, and 9 In all (a), (b) and (c), we find that the theme, No. 5, and No. 9 clump together; especially in (a) and (b). No. 2, No. 4 and No. 6 also clump together. No. 5 and No. 9 are contrapuntal variations of the theme, and their rhythmic structures are rather close together. In our experiment, we extracted salient pitch events by performing a time-span analysis, so that these three trees resembled each other.

No. 8 Although it has a similar rhythmic structure to the theme, No. 8 is in c-minor. In experiment (b), No. 8 was near the theme for this reason. In experiment (a), however, we could adequately distinguish the key by the pitch sensitivity.

No. 2, 4, and 6 No. 2, No. 4, and No. 6 include salient pitch events in the bass voice and thus are far from other variations. Those which consist of pitch events in the soprano voice tend to form a common tree, which reflects the original contour of the theme and thus form a macroscopic clump. In contrast, the monophonic representations of No. 2, No. 4, and No. 6 include an arpeggio of the harmony, so that the consonant notes tend to remain significant.

No. 3 No. 3 stays far from the clump of the theme because the chord progression is different.

No. 10 As we mentioned above, we excluded Nos. 10 - 12 from Figure 5. The monophonic representation of No. 10 is a mixture of two voices and its grouping structure in bar 3 is quite different from the other variations;

No. 12 No. 12 is in the triple meter, so that the distance easily tends to be larger. If we do compare it with others in our settings, we need to normalize the meter.

5. CONCLUSION

We extended GTTM with a preference rule for the pitch difference; that is, the important note in the local key is salient. According to this new rule, we revised the formula for the distance and calculated the distance in variations of Mozart K.265/300e. We showed that the time-span tree with pitch information adequately reflected the human cognitive perceptions of music, because the tree distance had the expected correlation with psychological similarity.

Our framework suggests the following issues. First, in general, variations are classified as follows [18]:

- decorative variation of melody with dissonant notes (No. 1, 3, and 7)
- rhythmic variation of melody (No. 1, 3, and 7)
- rhythmic variation of accompaniment (No. 2, 4, and 6)

- key changes (No. 8)
- harmonic variation (No. 2, 3, 4, 7, 10, and 11)
- contrapuntal variation (No. 5, 9, and 11)
- metrical variation (No. 12)
- exchanging melody and accompaniment (None in this piece)

It would be worth investigating if this normative classification correlates with the results of the structural analysis.

Second, the examinees may have been rather conscious of the rhythmic structure (Figure 5 (c)). We need to verify if this result was biased by our examinees or reflects a general tendency, by examining the differences in the musical experience of examinees.

Third, we put all the original pieces in a monophonic representation. Since the pitch information strongly depends on the chord, we must verify the adequacy of the obtained chord sequence; this implies if we claim the time-span tree reflects a cognitive reality, we need to treat a homophonic representation of music, and this will be our future work.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers 23500145 and 25330434). We thank K. Miyashita for help in the harmonic analysis and K. Okada for help in the statistical analysis.

6. REFERENCES

- [1] Bernabeu, J. F., Calera-Rubio, J., Iesta, J. M. and Rizo, D.: Melodic Identification Using Probabilistic Tree Automata, *Journal of New Music Research*, Vol. 40, Iss. 2, 2011.
- [2] ESCOM: 2007 Discussion Forum 4A. Similarity Perception in Listening to Music. *Musicae Scientiæ*
- [3] ESCOM: 2009 Discussion Forum 4B. Musical Similarity. *Musicae Scientiæ*
- [4] Gilbert, E. and Conklin, D.: A Probabilistic Context-Free Grammar for Melodic Reduction, *International Workshop on Artificial Intelligence and Music*, IJCAI-07, 2007.
- [5] Grachten, M., Arcos, J.-L. and de Mantaras, R.L.: Melody retrieval using the Implication/Realization model. 2005 MIREX. <http://www.music-ir.org/evaluation/mirexresults/articles/similarity/grachten.pdf>
- [6] Hamanaka, M., Hirata, K., Tojo, S.: Implementing “A Generative Theory of Tonal Music”. *Journal of New Music Research*, Vol. 35, Iss. 4, pp. 249–277 2007.
- [7] Hamanaka, M., Hirata, K. and Tojo, S.: Melody Morphing Method Based on GTTM, *Proceedings of ICMC 2008*, pp.155–158, 2008.
- [8] Hirata, K., Tojo, S. and Hamanaka, M.: Melodic Morphing Algorithm in Formalism, *LNAI6726*, Springer, pp. 338–341, 2011.
- [9] Hirata, K., Tojo, S. and Hamanaka, M.: Cognitive Similarity grounded by tree distance from the analysis of K.265/300e, *Proceedings of CMMR 2013*, pp. 415–430, 2013.
- [10] Kirilin., P. B.: *A Probabilistic Model of Hierarchical Music Analysis*, PhD thesis, University of Massachusetts Amherst, 2014.
- [11] F. Lerdahl and R. Jackendoff: *A Generative Theory of Tonal Music*, The MIT Press, Cambridge, 1983.
- [12] Lerdahl, F.: *Tonal Pitch Space*, Oxford University Press, 2001.
- [13] Schenker, H. (Oster, E. (trans.)) *Free Composition*, Longman, 1979. Original: *Der Freie Satz*, 1935.
- [14] Marsden, A.: Generative Structural Representation of Tonal Music, *Journal of New Music Research*, Vol. 34, Iss. 4, pp. 409–428, 2005.
- [15] Pampalk, E.: Computational Models of Music Similarity and their Application in Music Information Retrieval, PhD Thesis, Vienna University of Technology, 2006.
- [16] Rizo-Valero, D.: Symbolic Music Comparison with Tree Data Structure, Ph.D. Thesis, Universitat d’Alacant, Departamento de Lenguajes y Sistemas Informáticos, 2010.
- [17] Schedl, M., Knees, P. and Böck, S.: Investigating the Similarity Space of Music Artists on the Micro-Blogosphere, *Proceedings of ISMIR 2011*, pp. 323–328, 2011.
- [18] Randel, D., M.: *The new Harvard dictionary of music*, Harvard University Press, 1986.
- [19] Tojo S., and Hirata, K.: Structural Similarity Based on Time-span Tree, *Proceedings of CMMR 2012*, pp. 645–660, 2012.
- [20] Torgerson, W. S.: *Theory & Methods of Scaling*, New York: Wiley, 1958.

Appendix

Table 3 shows computationally calculated tree distance and psychological resemblance, as described in [9]. If an examinee, for instance, listens to Theme and variation No. 1 in this order, the ranking made by an examinee is listed in the first-row second-column cell (-0.73). The values in (b) are the averages over all the examinees.

Table 3. Computationally calculated tree distances and psychological resemblances (described in [9])

(a) Tree Distance without pitch information

	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12
Theme	183	177	195	183	117	249	162	15	21	363	262.5	246
No.1	-	228	332	326	264	360	219	174	204	456	409.5	421
No.2	-	-	264	216	246	282	105	168	186	438	391.5	423
No.3	-	-	-	252	262	320	259	188	198	462	334.5	379
No.4	-	-	-	-	238	246	213	176	186	424	387.5	399
No.5	-	-	-	-	-	276	243	114	108	414	298.5	325
No.6	-	-	-	-	-	-	291	234	264	378	409.5	449
No.7	-	-	-	-	-	-	-	153	171	429	376.5	400
No.8	-	-	-	-	-	-	-	-	30	348	259.4	255
No.9	-	-	-	-	-	-	-	-	-	378	277.5	261
No.10	-	-	-	-	-	-	-	-	-	-	406.5	403
No.11	-	-	-	-	-	-	-	-	-	-	-	298.5

(b) Average rankings by human listeners (listening in row→column order). Each listener rated their subjective similarity between two pieces in five grades: $\{-2, -1, 0, 1, 2\}$.

	Theme	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	No.11	No.12
Theme	-	-0.73	-0.91	-1.09	-0.82	1.18	-1.00	-1.45	-0.64	1.36	0.64	0.73	1.00
No.1	-1.00	-	-0.82	-0.73	-0.91	-0.64	0.36	-0.64	-1.45	-0.82	-0.82	-1.00	-0.64
No.2	-0.91	-0.36	-	-0.64	-0.27	-0.82	-0.45	-0.55	-1.55	-0.91	-0.09	-0.64	-0.91
No.3	-0.82	-0.45	-0.82	-	0	-0.91	-1.00	-0.36	-1.36	-0.73	-0.64	-0.73	-0.91
No.4	-1.00	-0.82	-0.73	0.18	-	-0.73	-0.82	-0.82	-1.73	-0.91	-0.45	-1.27	-1.00
No.5	1.27	-1.18	-0.91	-0.91	-0.64	-	-0.82	-1.09	-1.00	0.73	0.55	0.36	0.73
No.6	-1.18	0.27	-0.27	-0.45	-0.82	-0.64	-	-0.36	-1.64	-0.91	-0.55	-0.64	-0.91
No.7	-1.18	-0.64	-0.45	-0.18	-0.82	-0.73	-0.64	-	-1.18	-0.73	-0.36	-0.64	-0.73
No.8	-0.73	-1.27	-1.36	-1.55	-1.27	-0.73	-1.00	-1.36	-	-0.09	-1.09	-0.64	-0.91
No.9	1.27	-0.91	-0.91	-0.73	-1.09	0.91	-1.27	-0.82	-0.18	-	0.55	0.45	1.00
No.10	0.55	-0.82	-0.27	-0.64	-0.36	0.73	-0.45	-0.82	-1.00	0.73	-	0.18	0.45
No.11	0.64	-0.82	-0.91	-0.73	-0.91	0.55	-0.91	-1.09	-0.73	0.64	0.27	-	1.00
No.12	1.09	-1.18	-1.09	-1.00	-1.00	0.91	-1.00	-1.18	-0.91	1.09	0.36	0.82	-