Musical Rendering Models by Sequential Tension Rules

Tetsuya Mizutani and Yuki Nakata
University of Tsukuba, Tsukuba, Japan
Email: mizutani@cs.tsukuba.ac.jp

***Abstract*—Music informatics, especially study of musical rendition models is one of the most important areas in computer science and artificial intelligence. In order to investigate and develop musical rendition models, musical expressions of piano performances using hierarchical tonal tension rules have been analyzed by the authors. The rules are very complicated and hence it is hard to analyze large amount of music pieces. In this article musical expression is analyzed using sequential tension rules which are very simpler than the hierarchical ones. The results of the analyses by these two kinds of rules are also compared. From these comparisons, some notable results are obtained, i.e. though the results of the analyses in hierarchical rules are more precise than ones in sequential ones, the results in the latter are also useful to construct musical rendition models.**

***Index Terms*—tonal music theory, sequential and hierarchical tonal tension rules, multiple regression analysis**

I. Introduction

Music informatics is one of the most important areas in computer science, and also artificial intelligence [1] and [2]. Especially, investigation of musical rendition models is the most noteworthy. There are many researches for this area, and most of all research are based on the very pioneer music theory called *a Generative Theory of Tonal Music* (GTTM) [3] that is a cognitive theory of tonal music analyzing music by some strong hypotheses and structural rules of music, and introducing the hierarchical musical structures. *Tonal Pitch Space* (TPS) [4] is a succession of the theory that introduces structures of pitch, tone, chord, etc. It also introduces the *tonal tension structure* of chord progression, by which the tension value is defined. Moreover, the *melodic attraction* is also introduced by TPS for each pair of pitches indicating how stable, or natural, when a note of some pitch is performed after the other note of another pitch. The attraction value for each adjacent pair of pitches is defined as some kind of ratio between these strength values.

In order to create musical rendition models, musical expressions of piano performances using hierarchical tonal tension rules have been analyzed by the authors of this article [5] and [6]. The rules are very complicated since they must use the very complicated "*prolongational reduction tree*" which is the result of the structural and chordal analysis of each music pieces. Hence, it is hard to analyze large amount of music pieces using these rules.

In this article, musical expression is analyzed using sequential tension rules which are very simpler than the hierarchical ones, as reported in [7]. Also, the analysis results by these two kinds of rules are compared. From these analyses and comparison, some significant results are obtained, i.e. though the results of the analysis in hierarchical rules is more precise than ones in sequential ones, the results in the latter are useful to construct musical rendition models.

Manuscript received September 5, 2019; revised December 24, 2019.

II. Tonal Tension and Attraction

In this section, the harmonic tension and melodic attraction are briefly explained along [2].

A. Prolongational Reduction Tree

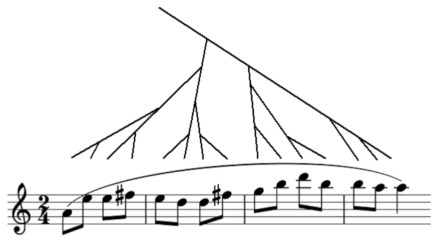

Figure 1. A prolongational reduction tree.

The *prolongational reduction structure* provides a cognition of tension-relaxation structure of music. In this structure, the hierarchy is concerned with stability expressed in terms of continuity and progression, and also tension or relaxation. The *prolongational reduction tree* represents this structure with some important musical notions, e.g. cadence structures, structural dominant, etc. as a binary tree. Fig. 1 shows an example of a prolongational reduction tree of a motif of "Orpheus in the Underworld" by J. Offenbach. In order to obtain such trees from music score, high level musical knowledge and deep understandings of GTTM and TPS must be required. In [8], there are some advance researches for the structure.

B. Harmonic Tension

The *chord distance* between two chords expresses the distance of them in the musical theory. In these definitions, an *event* of chord means the set of notes which sound simultaneously.

***Definition* 1.** The *chord distance* $\delta(x \rightarrow y)$ between two events of chords x and y is defined as

doi: 10.12720/jait.11.1.21-25

$$\delta(x \to y) = i + j + k \quad (1)$$

where i is the distance in the cycle of fifth of the diatonic scale of x and that of y; j the distance of two chords in the cycle of fifth; k the weighted sum of y minus that of x. The weighted sum of a chord is calculated as the following weight for each note in the chord: weight 4 if it is the root note, 3 if the fifth note in the chord, 1 if the other note in the chord, and 0 otherwise, i.e. it does not consist of the chord.

The *surface tension value* for each chord event is obtained from the event information of itself, defined by the notes that consist of the chord.

***Definition* 2.** The *surface tension value* $T_{sur}(y)$ for a chord event y is defined as

$$T_{sur}(y) = C_{sd} + C_{inv} + C_{nt} \quad (2)$$

where C_{sd} (scale degree) is 1 if the melodic note is the third or the fifth of y and 0 otherwise; C_{inv} (inversion) 2 if it is inversion, and 0 otherwise; C_{nt} (nonharmonic tone) 1 for seventh, 3 for diatonic nonharmonic, 4 for chromatic nonharmonic tone and 0 otherwise.

The *sequential tension value* for each chord event is obtained from the relationship between itself and the preceding chord.

***Definition* 3.** The *sequential tension value* $T_{seq}(y)$ for an event y is defined as

$$T_{seq}(y) = \delta(y_{prec} \to y) \quad (3)$$

where y_{prec} is the event that immediately precedes y in the sequence of events.

The *hierarchical tension* value for each chord event in the music score is defined by the prolongational reduction tree and the chord distance. It represents the repetition of tension and relaxation, and that of continuity and progression.

***Definition* 4.** The *inheritance tension value* $T_{inh}(y)$ for an event y is defined as

$$T_{inh}(y) = \delta(y_{root} \to y_1) + \cdots + \delta(y_n \to y) \quad (4)$$

where y_{root} is a root event y of a prolongational reduction tree, y_1 an event immediately inherited by y_{root}, y_{i+1} $(0 \le i < n)$ immediately inherited by y_i and y_n inherits the target event y immediately.

***Definition* 5.** The *hierarchical tension value* $T_{hier}(y)$ for an event y is defined as

$$T_{hier}(y) = \delta(y_{par} \to y) + T_{inh}(y_{par}) \quad (5)$$

where y_{par} is the parent chord of event in the prolongational reduction tree.

From these definitions, the harmonic hierarchical and sequential tension values are defined, each of which is obtained by adding surface tension value and hierarchical or sequential ones, respectively.

***Definition* 6.** The *harmonic hierarchical tension value* $T_{harhier}(y)$ and the *harmonic sequential tension value* $T_{harseq}(y)$ are defined as

$$T_{harhier}(y) = T_{hier}(y) + T_{sur}(y), \text{ and} \quad (6)$$

$$T_{harseq}(y) = T_{seq}(y) + T_{sur}(y) \quad (7)$$

C. Melodic Attraction

The melodic attraction represents a degree how natural one pitch transits another (succeeding) one. The realized voice-leading attraction expresses how much the succeeding chord attracts from the original one. These values are obtained from the anchoring strength which expresses how stable the target pitch in the chord.

***Definition* 7.** The *anchoring strength* is defined for each pitch in the context of chord as follows: value 4 if the pitch is root note in the chord, 3 if the third or the fifth, 2 if it is other pitch belonging in the chord and 1 otherwise, i.e. it does not consist of the chord.

***Definition* 8.** The *melodic attraction* $\alpha(p_1 \to p_2)$ for a pitch p_1 to another p_2 with $p_1 \neq p_2$ is defined as

$$\alpha(p_1 \to p_2) = \frac{s_2}{s_1} \times \frac{1}{n^2} \quad (8)$$

where s_1 and s_2 are the anchoring strength of p_1 and p_2, respectively, and n is the number of semitone interval between p_1 and p_2.

***Definition* 9.** The *realized voice-leading attraction* $\alpha_{rvl}(C_1 \to C_2)$ for a chord C_1 to another chord C_2 such that not all pitches are identical is defined as

$$\alpha_{rvl}(C_1 \to C_2) = \alpha_{r1} + \cdots + \alpha_{rn} \quad (9)$$

where $\alpha_{r1}, \ldots, \alpha_{rn}$ are the melodic attractions for all the voices in C_1 to C_2.

III. Experiments

A. Musical Data

Score and performance data used in this article are obtained from CrestMusePEDB [9] provided by the CrestMuse Project. The titles of the music pieces and the composers with the pianists played these music pieces using as data in the experiments are shown in Table I.

TABLE I. MUSICAL DATA

Composer	Music Title	Pianist
L. v. Beethoven	Piano Sonata No. 8 in C minor, Op. 13, “Sonata Pathétique”	C.Arrau
		V. D. Ashkenazy
		W. Backhaus
F.F. Chopin	Nocturne in E-flat major, Op. 9, No. 2	V.D. Ashkenazy
		S. S. Bunin
		V. S. Horowitz
		M. J. Pires
M.A. Mozart	Piano Sonata No. 11 in A major, K. 331, Movement 1	C. Eschenbach
		G. H. Gould
		I. Haebler
		L. Kraus
		A. deLarrocha
		H. Nakamura
		M. J. Pires
		N. Shimizu
	Piano Sonata No. 16 in C major, K. 545, Movement 1	M. J. Pires

B. Multiple Regression Analysis

The *multiple regression analysis* [10] with the *Akaike Information Criterion* (AIC) [11] are used in the analyses.

The multiple regression analysis is a statistical technique for estimating the relationship among a dependent variable and some independent variables. AIC is used as a measure of the relative quality of statistical models. *The overfitting* may happen if the independent variables are too numerous. To avoid this, some independent variables must be selected and others are rejected appropriately. AIC introduces one optimum solution for such problems.

The *p-value* indicates the probability that the partial regression coefficient is 0. In general, it is the strong evidence against null hypothesis, i.e. the coefficient is meaningful, if the value is less than 0.05, when the significant level is 5%.

The *one sample t-test* determines whether two sets of data are significantly different from each other or not. In this article, it is used to test whether the null hypothesis is true for the average of the partial regression coefficients of independent variables obtained from the multiple regression analysis. The hypothesis implies that the two samples are independent, and vice versa.

C. Experiments

In order to calculate the chord distance in the scores used in the calculations of the tension values, each event of chord must be identified as the correspondent chord names. In the identification, high level musical sense and complicated interpretation knowledge for chords are necessary, and moreover, the results of identification may not be unique. Therefore, the chord analysis and interpretation must be formalized in a regular way. For this purpose, *music21 class library* [12] of *Python* [13] is used in the experiments.

The value of the dependent variable is local tempo or volume of the corresponding event, each of which is calculated individually. The values of independent variables are the harmonic hierarchical or sequential tension values with the realized voice-leading attraction values of each event, whose values are nonnegative integers, and additionally, the existence and positions of each expression marks, e.g. start point of accent, inside of crescendo, endpoint of decrescendo, etc., whose values are 1 if exists and 0 otherwise.

Table II-Table V show the results of the regression analyses for "Sonata Pathétique", "Nocturne", "K.331" and "K.545", respectively. In these tables, "Indep" means the independent variables used in the analyses, and "Coeff" is the coefficient of determination calculated by the analysis. In the Tension column, "hie" and "seq" indicate that the analysis is used hierarchical or sequential tension value, respectively, as one of the independent variables. In the AIC column, "Y" indicates that the independent variable of tension is selected by AIC, while "N" does it is rejected. In the "p" column, "**" indicates that the coefficient of tension value is meaningful under the significant level 0.05 i.e. the corresponding p-value are less than 0.05, and also, "*" does that the significant level is 0.1, whereas blank means it is not significant and "-" no data (= the tension value is rejected by AIC).

It must be noted that the coefficient of determination of the hierarchical tension value of volume of Gould and that of the sequential one of volume of Haebler in Table IV are not obtained from the analysis since every independent variable other than the interrupt is rejected by AIC.

From Table II-Table V, the following facts are obtained.

(1) The tension values vary widely depending on music pieces and pianists.
(2) The values of the coefficient of determination of the hierarchical tension values are better than those of the sequential ones. Thus, the former can describe music expression more precisely than the latter.
(3) In some performances, e.g. in Nocturne by Horowitz, the sequential tension rules can be described these expression trends more precisely than the hierarchical ones.

Table VI and Table VII show the averages of the coefficient of determination of tempo and volume, respectively, for each music pieces.

From the two tables, some results are obtained as follows.

(1) For tempo, the averages of two coefficients are rather similar. It means that a sequential tension structures are useful for creating a musical rendition model.
(2) For volume, the average of the coefficient used the hierarchical tension structures are better than that used the sequential ones.
(3) Nonetheless, most coefficient values are rather small and vary widely.

Table VIII is the list of independent variables whose coefficients are meaningful from the results of the t-test. In this table, "_M" after the names of the independent variable means the corresponding independent variable is for the melody part, while "_A" for the accompaniment. The last row "tension" indicates the sequential tension.

This table indicates the followings.

(1) Some tendency of the musical expression can be obtained, at least qualitatively. Namely, these independent variables can explain some trends of each music expression.
(2) Some curious properties, which are already reported in [6], are also obtained. In especial, "piano", which describes to play softly, affects not volume but tempo.
(3) All of the experiments use "attraction" as one of the independent variables, however, it does not obtain in this table. It means that attraction is rather not meaningful for constructing musical rendition models.

From these experiments, it can be concluded as follows. The coefficients obtained from the results of these experiments using sequential rules are rather close to them using hierarchical ones. Namely, although the results of the analyses in hierarchical rules is more precise than ones in sequential ones, the sequential rules are useful to develop musical rendition models.

TABLE II. RESULT OF THE ANALYSIS OF SONATA PATHÉTIQUE

Pianist	Indep	Tension	Coeff	AIC	p
Arrau	tempo	hie.	0.157	Y	**
		seq.	0.176	Y	**
	volume	hie.	0.284	Y	**
		seq.	0.319	Y	**
Ashkenazy	tempo	hie.	0.158	N	-
		seq.	0.158	N	-
	volume	hie.	0.257	Y	**
		seq.	0.0933	N	-
Backhaus	tempo	hie.	0.207	N	-
		seq.	0.248	Y	
	volume	hie.	0.329	Y	**
		seq.	0.221	N	-

TABLE III. RESULT OF THE ANALYSIS OF NOCTURNE

Pianist	Indep	Tension	Coeff	AIC	p
Ashkenazy	tempo	hie.	0.245	N	-
		seq.	0.245	N	-
	volume	hie.	0.471	Y	**
		seq.	0.369	Y	*
Bunin	tempo	hie.	0.252	N	-
		seq.	0.252	N	-
	volume	hie.	0.346	Y	
		seq.	0.330	N	-
Horowitz	tempo	hie.	0.172	N	-
		seq.	0.172	N	-
	volume	hie.	0.441	N	-
		seq.	0.457	Y	**
Pires	tempo	hie.	0.281	N	-
		seq.	0.281	N	-
	volume	hie.	0.540	Y	
		seq.	0.534	N	-

TABLE IV. RESULT OF THE ANALYSIS OF K.331

Pianist	Indep	Tension	Coeff	AIC	p
Eschenbach	tempo	hie.	0.355	Y	
		seq.	0.333	N	-
	volume	hie.	0.233	Y	**
		seq.	0.023	Y	*
Gould	tempo	hie.	0.0868	N	-
		seq.	0.136	Y	*
	volume	hie.	-	-	-
		seq.	0.130	Y	**
Haebler	tempo	hie.	0.688	N	-
		seq.	0.688	N	-
	volume	hie.	0.314	Y	*
		seq.	-	-	-
Kraus	tempo	hie.	0.428	N	-
		seq.	0.425	Y	
	volume	hie.	0.582	Y	**
		seq.	0.200	Y	*
deLarrocha	tempo	hie.	0.451	N	-
		seq.	0.451	N	-
	volume	hie.	0.557	Y	**
		seq.	0.326	Y	
Nakamura	tempo	hie.	0.0803	N	-
		seq.	0.0803	N	-
	volume	hie.	0.0336	N	-
		seq.	0.0336	N	-
Pires	tempo	hie.	0.286	N	-
		seq.	0.286	N	-
	volume	hie.	0.474	Y	**
		seq.	0.0971	Y	
Shimizu	tempo	hie.	0.417	N	-
		seq.	0.417	N	-
	volume	hie.	0.710	Y	**
		seq.	0.542	Y	**

TABLE V. RESULT OF THE ANALYSIS OF K.545

Pianist	Indep	Tension	Coeff	AIC	p
Pires	tempo	hie.	0.339	Y	**
		seq.	0.236	Y	**
	volume	hie.	0.383	Y	**
		seq.	0.274	Y	*

TABLE VI. COEFFICIENT OF DETERMINATION OF TEMPO

MUSIC	Tension	
	Hierarchical	Sequential
Sonata Pathétique	0.232	0.208
Nocturne in E-flat major	0.238	0.238
K.331	0.349	0.352
K.545	0.355	0.333
Average	0.293	0.283

TABLE VII. COEFFICIENT OF DETERMINATION OF VOLUME

Music	Tension	
	Hierarchical	Sequential
Sonata Pathétique	0.290	0.211
Nocturne in E-flat major	0.450	0.423
K.331	0.400	0.215
K.545	0.233	0.023
Average	0.343	0.218

TABLE VIII. INDEPENDENT VARIABLES MEANINGFUL FOR RENDITION

	tempo		volume	
Indep	Coeff	p	Coff	p
Intercept	57.4	**	37.3	**
piano_start_M	-13.3	*		
forte_start_M			13.6	**
slur_start_M	5.17	**		
slur_cont_M	9.41	*		
slur_end_M	-3.76	**		
cre_cont_M			8.38	**
decre_cont_M			-9.97	**
staccato_M	-9.22	**		
accent_M	9.86	**		
slur_cont_A			5.88	*
cre_start_A			14.6	**
cre_cont_A			12	*
pedal_on			4.58	**
pedal_off	-6.59	**	-3.8	**
tension			4.5	**

IV. SUMMARY

The analyses of musical expression using sequential and hierarchical tension values as one of the independent variables are executed. The results show that the sequential tension value is as useful as the hierarchical ones for developing musical rendition models and also program systems based on these models that creates artistic performances automatically, although the hierarchical rules can describe features of musical expression more precisely than sequential rules. Though some tendency or trends of the musical expression are obtained from these experiments, it is hard to decide rendition models from only small numbers of samples using this article. Therefore, the authors of this articles must investigate massive data of music performances. In order to realize it, the authors are investigating some methods to develop musical data automatically from existing audio data of music pieces such as CDs, some kinds of music performance data that can be obtained from the internet, and so on.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Mizutani conducted the research; Nakata analyzed the data; Mizutani wrote the paper; all authors had approved the final version.

ACKNOWLEDGMENT

The authors wish to thank Yamato Iwami and Kei Hasegawa for providing the authors with a lot of fruitful knowledge on music, statistics and program library for musical analyses.

REFERENCES

[1] M. Müller, *Information Retrieval for Music and Motion*, Springer-Verlag Berlin Heidelberg, 2007.

[2] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, “Automatic music transcription: An overview,” *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20-30, 2019.

[3] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*, Cambridge: The MIT Press, 1983.

[4] F. Lerdahl, *Tonal Pitch Space*, Oxford: Oxford University Press, 2001.

[5] T. Mizutani and Y. Iwami, “Analysis of musical expression using tonal tension and attraction models,” in *Proc. the 2nd International Conference on Applied Mathematics, Simulation and Modelling*, 2017, pp. 344-348.

[6] T. Mizutani, S. Nakahara, and T. Takeda, “Comparison of musical expression of distinguished and unnamed pianists in tonal tension and attraction models,” in *Proc. the 3rd International Conference on Control, Automation and Artificial Intelligence*, 2018, vol. 1, pp. 162-165.

[7] Y. Nakata, “Analysis of music expression using sequential tonal tension models (in Japanese),” graduation thesis, College of Information Science, University of Tsukuba, 2019.

[8] M. Hamanaka, K. Hirata, and S. Tojo, “GTTM database and manual time-span tree generation tool,” in *Proc. the 15th Sound and Music Computing Conference*, 2018, pp. 462-467.

[9] CrestMusePEDB. [Online]. Available: http://www.crestmuse.jp/pedb/

[10] A. Sen and M. Srivastava, *Regression Analysis - Theory, Methods, and Applications*, 4th ed., Berlin: Springer-Verlag, 2011.

[11] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in *Breakthroughs in Statistics: Foundations and Basic Theory*, S. Kotz and N. L. Johnson, Eds., Springer-Verlag, 1992, vol. 1, pp. 610-624.

[12] M. Cuthbert. Music21: A Toolkit for Computer-Aided Musicology. [Online]. Available: http://web.mit.edu/music21/

[13] Python. [Online]. Available: https://www.python.org/

Tetsuya Mizutani was born in Japan in 1960. He received his B.Eng., M.Eng. and Ph.D. in Eng. from University of Tsukuba in 1983, 1985, and 1988, respectively. He is currently a fulltime assistant professor of Department of Information Engineering at University of Tsukuba. His research interests are musical informatics and program theory, especially theoretical foundation of musical information and also formal verification and analysis of real time intellectual program systems.

Yuki Nakata was born in Japan in 1996. He received his B.Eng. from College of Information Science, University of Tsukuba in 2019. His research interest is computer science, especially music informatics. Mr. Nakata is currently pursuing a system engineer at Nova System Co., LTD., Japan.