

FATTA: FULL AUTOMATIC TIME-SPAN TREE ANALYZER

Masatoshi Hamanaka

University of Tsukuba

hamanaka@iit.tsukuba.ac.jp

Keiji Hirata

NTT Communication Science
Laboratories

hirata@brl.ntt.co.jp

Satoshi Tojo

Japan Advanced Institute of
Science and Technology

tojo@jaist.ac.jp

ABSTRACT

We developed a music analysis system called a full automatic time-span tree analyzer (FATTA), which analyzes a piece of music based on the generative theory of tonal music (GTTM). We previously developed an automatic time-span tree analyzer (ATTA), which can acquire a time-span tree of GTTM. Although the ATTA has adjustable parameters for controlling the weight or the priority of each rule, these parameters have to be set manually. This takes a long time because finding the optimal values of the settings themselves takes a long time. The FATTA can automatically estimate the optimal parameters by introducing a feedback loop from higher-level structures to lower-level structures based on the stability of the time-span tree. Experimental results showed that the performance of FATTA outperformed a baseline performance of the ATTA. We hope to distribute the time-span tree as the content for various musical tasks, such as searching and arranging music.

1. INTRODUCTION

We have been constructing music analysis system based on the music theory of Generative Theory of Tonal Music (GTTM) by Lerdahl and Jackendoff [1]. Musical theory provides us with the methodologies for analyzing and transcribing musical knowledge, experiences, and skills from a musician's way of thinking. Our concern is whether or not the concepts necessary for music analysis are sufficiently externalized in musical theory. We consider the GTTM to be the most promising theory among the many that have been proposed [2–4], in terms of its ability to formalize musical knowledge, because the GTTM captures the aspects of the musical phenomena based on the Gestalt occurring in music and is presented with relatively rigid rules.

The GTTM is composed of four modules, each of which assigns a separate structural description to a listener's understanding of a piece of music. These four modules output a grouping structure, metrical structure, time-span tree, and prolongational tree, respectively.

Our goal is to automatically derive a time-span tree that assigns a hierarchy of 'structural importance' to the notes of a piece of music. The time-span tree provides a summarization of a piece of music, which can be used as the representation of a search, by analyzing the results from the GTTM, resulting in a music retrieval system [5]. It can also be used for performance rendering [6–8] and reproducing music [9]. These

systems enable users to manipulate music using a time-span tree, disregarding the surface structure of the music. However, the time-span trees in these systems need to be manually analyzed by experts in musicology. Although a computer model of the GTTM [10] is capable of producing a time-span tree, we, humans, need to choose applicable rules at each stage of the processing. We have previously constructed an automatic time-span tree analyzer (ATTA), which can derive grouping structure, metrical structure, and a time-span tree based on an extended GTTM (exGTTM) that we proposed. The exGTTM re-formalizes the rules of the GTTM and establishes an algorithm for acquiring a time-span tree [11–13] (Figure 1).

However, the ATTA has a problem properly analyzing results: To analyze them properly, we have to manually tweak the parameters of the ATTA on a computer screen, and such parameter tuning is tedious and laborious. To overcome this problem, we have developed a method to automatically estimate the optimal parameters using a feedback loop from the higher-level structures to the lower-level structures based on the stability of the time-span tree.

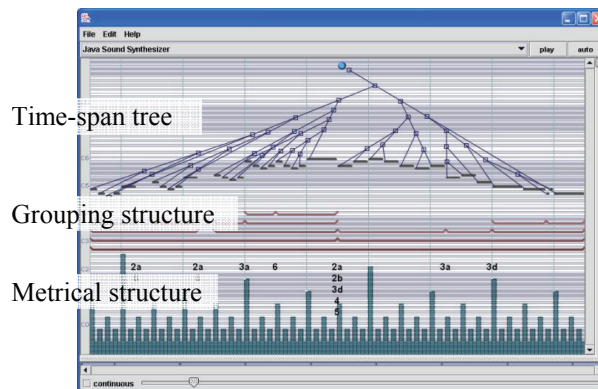


Figure 1. ATTA: automatic time-span tree analyzer.

2. EXGTTM AND ATTA

To get a music theory to operate on a computer, however, we must overcome some widely recognized fundamental difficulties. One is giving an ambiguous concept a firm definition, and the other is supplementing the lack of necessary concepts (externalization). Here, note that we distinguish two kinds of ambiguity in music analysis: one involves the musical understanding of humans, and the other concerns the representation of a music theory. In our previous work [11], we extended the GTTM by full externalization and parameterization and proposed a machine-executable extension of the GTTM, exGTTM.

We implemented the exGTTM on a computer that we call ATTA (Figure 2). The externalization in mechanizing the GTTM includes introducing an algorithm for generating a hierarchical structure of the time-span tree in the mixed manner of top-down and bottom-up. Such an algorithm has not been represented for GTTM. The parameterization includes introducing a parameter controlling the priorities of rules to avoid a conflict among the rules as well as parameters for controlling the shape of the hierarchical time-span tree.

The significance of full externalization and parameterization is twofold: precise controllability, and coverage of the manual results. Whenever we find a correct result that the exGTTM cannot generate, we introduce new parameters for the exGTTM and give them proper values so that it can generate the correct result. In this way, we repeatedly externalize and introduce new parameters until we can obtain all of the results that humans consider correct. In total, we have introduced 15 parameters for grouping-structure analysis, 18 for metrical-structure analysis, and 13 for time-span reduction (Table 1).

We appropriately supply the missing parameters and make implicit parameters explicit. The parameters introduced by exGTTM are categorized into three categories: identified, implied, and unaware. For the first category, a parameter is identified in GTTM but is not assigned concrete values. Hence, we value such a parameter. For the second category, a parameter is implied in GTTM. Hence, we make it explicit. For the third category, we need to complement parameters that

are not recognized in the original theory, since some of them may nearly lack any musicological meaning.

3. FATTA

Although the ATTA can automatically acquire a time-span tree, because the parameters were only controlled manually, it took too much time to find a set of optimal parameters. Therefore, we developed a method to automatically estimate the optimal parameters of the ATTA.

In the GTTM [1], there are two rules that are not implemented on the ATTA: GPR7 and TSRPR5.

GPR7 (Time-Span and Prolongational Stability) prefer a grouping structure that results in more stable time-span and/or prolongational reductions.

TSRPR5 (Metrical Stability) In choosing the head of time-span T, prefer a choice that results in more stable choice of metrical structure.

These rules require information from later processes, such as time-span/prolongational reductions, to be sent back to the earlier processes.

To automatically estimate the optimal parameters, we have to evaluate the level of time-span tree stability that are derived by the ATTA. We use GPR7 and TSRPR5 for calculating the level of stability. Figure 2 shows the process flow of the FATTA, which consist of the ATTA and a feedback loop by the GPR7 and TSRPR5.

	Parameters	Description
Grouping structure	$S_{GPR\ R}$ ($0 \leq S_{GPR\ R} \leq 1$)	Strength of each grouping preference rule. The larger the value is, the stronger the rule acts. $R \in \{ 2a, 2b, 3a, 3b, 3c, 3d, 4, 5, \text{ and } 6 \}$
	σ ($0 \leq \sigma \leq 0.1$)	Standard deviation of a Gaussian distribution, the average of which is the boundary by GPR5. The larger the value is, the wider its skirt becomes.
	W_m ($0 \leq W_m \leq 1$)	Balance between temporal similarity of attack points and that of pitch difference in GPR6. The larger the value is, the more the system estimates the pitch difference.
	W_l ($0 \leq W_l \leq 1$)	Weight for the length of parallel phrases. The larger the values is, the more the length of parallel phrases is prioritized in GPR6.
	W_s ($0 \leq W_s \leq 1$)	Balance determining whether the note i becomes the ending note of a group or the beginning note of the following group in GPR6. The larger the value is, the more the note tends to be the ending note.
	T_{GPR4} ($0 \leq T_{GPR4} \leq 1$)	Threshold at which the effects of GPR2,3 are considered to be salient in GPR4. The smaller the value is, the more probably GPR4 is applied.
	T^{low} ($0 \leq T^{low} \leq 1$)	Threshold in the lower-level boundary. The smaller the value is, the more salient the boundary becomes.
Metrical structure	$S_{MPR\ R}$ ($0 \leq S_{MPR\ R} \leq 1$)	Strength of each metrical preference rule. The larger the value is, the stronger the rule acts. $R \in \{ 1, 2, 3, 4, 5a, 5b, 5c, 5d, 5e, \text{ and } 10 \}$
	W_m ($0 \leq W_m \leq 1$)	Balance between temporal similarity of attack points and that of pitch difference in MPR1. The larger the value is, the more the system estimates the pitch difference.
	W_l ($0 \leq W_l \leq 1$)	Weight for the length of parallel phrases. The larger the value is, the more the length of parallel phrases is prioritized in MPR1.
	W_s ($0 \leq W_s \leq 1$)	Balance determining whether the note i becomes the ending note of a group or the beginning note of the following group in MPR1. The larger the value is, the more the note tends to be the ending note.
	$T_{MPR\ R}$ ($0 \leq T_{MPR\ R} \leq 1$)	Value of the threshold that decides whether each rule is applicable. $R \in \{ 4, 5a, 5b, 5c \}$
Time-span tree	$S_{TSRPR\ R}$ ($0 \leq S_{TSRPR\ R} \leq 1$)	Strength of each rule. The larger the value is, the stronger the rule acts. $R \in \{ 1, 2, 3, 4, 5a, 5b, 5c, 5d, 5e, \text{ and } 10 \}$
	W_m ($0 \leq W_m \leq 1$)	The balance between the temporal similarity of attack points and that of the pitch difference in TSRPR4. The larger the value is, the more the system estimates the pitch difference.
	W_l ($0 \leq W_l \leq 1$)	The weight for the length of parallel phrases. The larger the values is, the more the length of parallel phrases is estimated in TSRPR4.
	W_s ($0 \leq W_s \leq 1$)	The balance determines whether the note i becomes the ending note of a group or the beginning note of the following group in TSRPR4. The larger the value is, the more the note tends to be the ending note.

Table 1. Adjustable parameters of the exGTTM and ATTA.

3.1. Implementation of GPR7 with Tonal Pitch Space

GPR7 is the rule applied to the feedback loop between the time-span/prolongational reduction and grouping structure analysis. The rule leads to a preference for a grouping structure that results in a more stable time-span and/or prolongational reductions.

D_{GPR7} indicates the holding level of GPR7, which varies continuously between 0 and 1. We define D_{GPR7} as

$$D_{GPR7} = \frac{\sum_i \text{distance}(p(i), s(i)) \times \text{size}(i)^2}{\sum_i \text{size}(i)^2} \quad (1)$$

where, i indicates the head of time-span, which has primary and secondary branches, denoted by $p(i)$ and $s(i)$, respectively. The $\text{distance}(x, y)$ indicates the distance between note x to note y in the tonality of the piece, which are defined in Tonal Pitch Space music theory by Lerdahl [14]. We normalized the distance from 0 to 1. The $\text{size}(i)$ indicates the length of time-span which has head i . When calculating D_{GPR7} , we use the square of $\text{size}(i)$ for weightings because of empirical reason.

3.2. Implementation of TSRPR5

TSRPR5 is the rule applied to the feedback loop between the time-span reduction and metrical structure analyzer. The rule leads to a preference that results in a more stable choice of metrical structure in choosing the head of a time-span. D_{TSRPR5} indicates the holding level of TSRPR5, which varies continuously between 0 and 1. We define D_{TSRPR5} as follows

$$D_{TSRPR5} = \frac{1}{\sum_i \text{size}(i)^2} \times \sum_i \begin{cases} \text{size}(i)^2 & \text{dot}(p(i)) \geq \text{dot}(s(i)) \\ 0 & \text{dot}(p(i)) < \text{dot}(s(i)) \end{cases} \quad (2)$$

where, $\text{dot}(x)$ indicates the number of metrical dots of note x .

3.3. Optimization of adjustable parameters

The optimal parameter sets of the ATTA can be obtained by maximizing the average of D_{GPR7} ($0 \leq D_{GPR7} \leq 1$) and D_{TSRPR5} ($0 \leq D_{TSRPR5} \leq 1$). Because there are 46 adjustable parameters, it requires a large amount of time to calculate all the combinations of parameter sets. In order to decrease the calculation time, we constructed the following algorithm:

- (1) Maximize the average of D_{GPR7} and D_{TSRPR5} by changing a parameter from minimum value to maximum value
- (2) Repeat (1) for all the parameters
- (3) Iterate (1) and (2), as long as the average of D_{GPR7} and D_{TSRPR5} is increased from the previous iteration.

4. EXPERIMENTAL RESULTS

We evaluated the performance of FATTA using an F-measure, which is given by the weighted harmonic mean of precision P (proportion of selected groupings/dots/heads that are correct) and recall R (Proportion of correct groupings/dots/heads that are identified). In calculating F-measure of the grouping analyzer and time-span tree analyzer, we did not take consider the possibility that a low-level error is propagated up to a higher level; we counted wrong answers without regard to the differences in grouping levels and time-span levels.

This evaluation required us to prepare accurate data for the grouping structure, metrical structure, and time-span tree. We collected hundred pieces of 8-bar-length, monophonic, classical music and asked people with expertise in musicology to analyze them

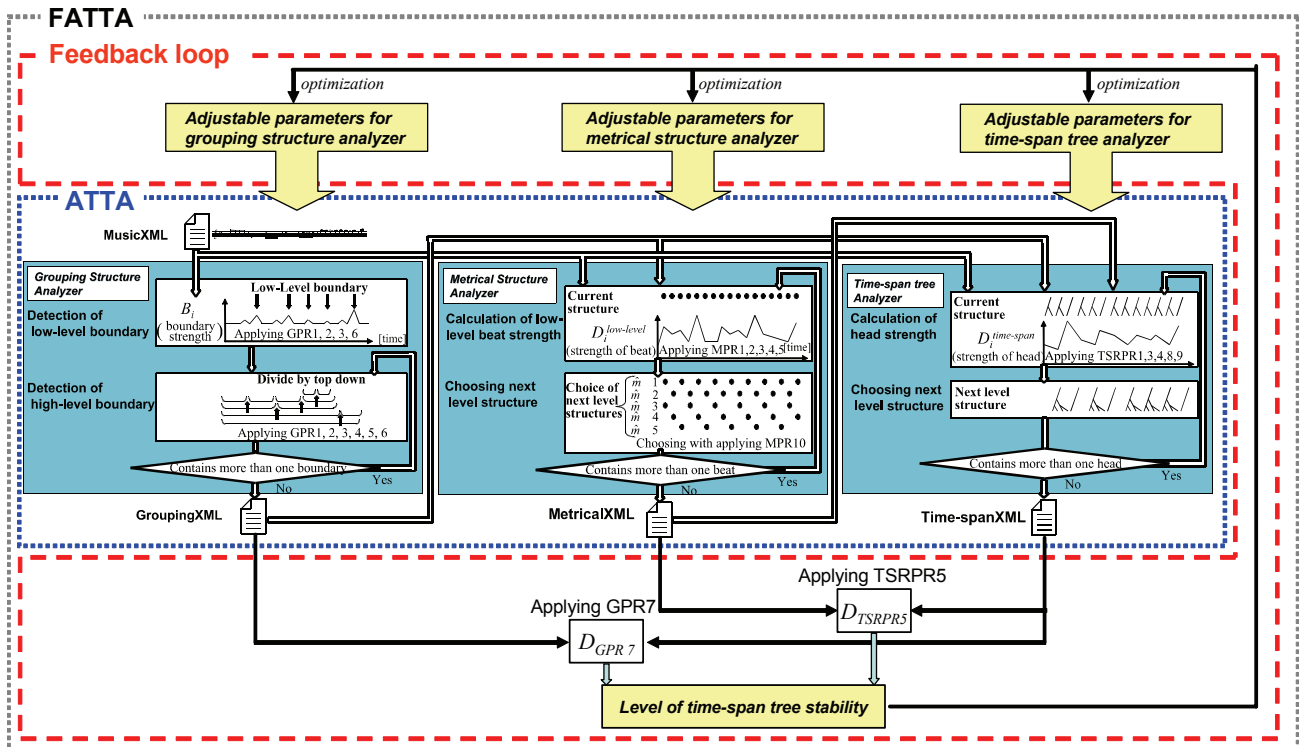


Figure 2. Processing flow of full automatic time-span tree analyzer (FATTA).

manually with faithful regard to the GTTM. These manually produced results were cross-checked by three other experts.

The grouping, metrical and time-span tree structures will change depending on the adjustable parameters. To evaluate the baseline performance of the ATTA, we used the following default parameters: $S_{rules} = 0.5$, $T_{rules} = 0.5$, $W_s = 0.5$, $W_r = 0.5$, $W_l = 0.5$, and $\sigma = 0.05$. The range of parameters of T_{rules} , W_s , W_r , W_l is 0 to 1.0 and resolution is 0.1. The range of a parameter of σ is 0 to 0.1 and resolution is 0.01.

After optimizing the parameter set by the FATTA, the average F-measure for grouping structure was 0.40, metrical structure is 0.88, and time-span tree is 0.35, which outperformed a baseline performance of the ATTA (Table 2).

5. CONCLUSION

We developed a music analyzing system called a FATTA, which derives the time-span tree of the GTTM without requiring manual optimization of adjustable parameters of an ATTA. To automatically estimate the optimal parameters, we use GPR7 and TSRPR5 for calculating the level of time-span tree stability. We developed an algorithm that enabled a decrease in this calculation time. As a result, the FATTA can optimize the parameters and can acquire grouping structure, metrical structure, and time-span tree easily. Our experimental results show that the performance of FATTA outperformed the baseline F-measure of ATTA.

REFERENCES

- [1] Lerdahl, F., and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Massachusetts, 1983.
- [2] Cooper, G., and Meyer, L. B. *The Rhythmic Structure of Music*. The University of Chicago Press, 1960.
- [3] Narmour, E. *The Analysis and Cognition of Basic Melodic Structure*. The University of Chicago Press, 1990.
- [4] Temperley, D. *The Cognition of Basic Musical Structures*. MIT press, Cambridge, 2001.

- [5] Hirata, K., and Matsuda, S. "Interactive Music Summarization based on Generative Theory of Tonal Music". *Journal of New Music Research*, 32:2, 165-177, 2003.
- [6] Todd, N. "A Model of Expressive Timing in Tonal Music". *Musical Perception*, 3:1, 33-58, 1985.
- [7] Widmer, G. "Understanding and Learning Musical Expression", Proceedings of International Computer Music Conference, pp. 268-275, 1993.
- [8] Hirata, K., and Hiraga, R. "Ha-Hi-Hun plays Chopin's Etude", Working Notes of IJCAI-03 Workshop on Methods for Automatic Music Performance and their Applications in a Public Rendering Contest, pp. 72-73, 2003.
- [9] Hirata, K., and Matsuda, S. "Annotated Music for Retrieval, Reproduction, and Sharing", Proceedings of International Computer Music Conference, pp. 584-587, 2004.
- [10] Nord, T. *Toward Theoretical Verification: Developing A Computer Model of Lerdahl and Jackendoff's Generative Theory of Tonal Music*, Dissertation, The University of Wisconsin, Madison, 1992.
- [11] Hamanaka, M., Hirata, K., and Tojo, S.: "ATTA: Automatic Time-span Tree Analyzer based on Extended GTTM", *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR2005)*, pp. 358-365, 2005.
- [12] Hamanaka, M., Hirata, K., and Tojo, S.: "Automatic Generation of Metrical Structure based on GTTM", *Proceedings of the 2005 International Computer Music conference (ICMC2005)*, pp. 53-56, 2005.
- [13] Hamanaka, M., Hirata, K., and Tojo, S.: "Automatic Generation of Grouping Structure based on the GTTM", *Proceedings of the 2004 International Computer Music conference (ICMC2004)*, pp. 141-144, 2004.
- [14] Lerdahl, F.: *Tonal Pitch Space*, Oxford University Press, 2001.

Melodies	Grouping Structure Analyzer		Metrical Structure Analyzer		Time-Span Tree Analyzer	
	Baseline	FATTA	Baseline	FATTA	Baseline	FATTA
1. Grande Valse Brillante	0.21	0.32	0.88	0.88	0.37	0.41
2. Moments Musicaux	0.24	0.60	0.95	1.00	0.58	0.74
3. TrukishMarch	0.67	0.67	0.91	0.96	0.68	0.80
4. Anitras Tanz	0.29	0.71	0.82	0.82	0.55	0.52
5. Valse du Petit Chien	0.04	0.28	0.87	0.95	0.17	0.57
	:	:	:	:	:	:
Total (100 melodies)	0.46	0.48	0.84	0.89	0.44	0.49

Table 2. F-measure for baseline performance of ATTA and FATTA.