Structural Analysis of Utterances during Guitar Instruction

Nami Iino[1,2], Hiroya Miura[2], Hideaki Takeda[1],
Masatoshi Hamanaka[2], and Takuichi Nishimura[3] ⋆

[1] National Institute of Informatics
[2] RIKEN Center for Advanced Intelligence Project
[3] Japan Advanced Institute of Science and Technology
nami-iino@nii.ac.jp

Abstract. The physical environments for musical instrument instruction include a mixture of various types of information such as performance sounds and speech. In our previous study, we analyzed the speech segments of audio data recorded in real one-on-one classical guitar lessons. In the current work, we annotate two types of labels for the teacher's utterance information and analyze them structurally by applying the Generative Theory of Tonal Music (GTTM) to summarize the lessons. Our findings revealed a commonality in the interpretation of utterance groupings and demonstrate that the labels for semantically categorizing the content of teachers' utterances are useful in determining the hierarchy.

1 Introduction

In musical instrument lessons, teachers and learners often record audio or video for later review and reflection. However, these private recordings are rarely made publicly available as a research resource. Additionally, the specific terms and instructional content utilized in real lessons have not been analyzed in detail. Therefore, we previously collected sound information from one-on-one classical guitar lessons to help clarify the features of the music teaching-learning process [1].

In the current study, we structurally analyze the utterances that occur during a guitar lesson with the aim of summarizing the overall lesson. Specifically, We annotated the teacher's utterances data with two types of labels and generated tree structures using an analysis method based on the Generative Theory of Tonal Music (GTTM) [2]. Through a comparison of the tree structures by two analysts, we examined the common structural patterns and rules for the aggregation of tree structures. Our findings revealed a commonality in the analysts' interpretations and showed that labels related to instructional content are useful in determining hierarchy.

⋆ This study was partially supported by Kayamori Foundation of Informational Science Advancement.

A research [3] proposed an interactive information structuring method for meeting minutes utilizing the GTTM. In that study, to explicitly express the intentions contained in discussions, a tree structure and a method for its extraction were implemented to represent hierarchical importance levels based on verbal and non-verbal information. Our study aims to find a method suitable for lessons utilizing this same approach. We also take advantage of CROCUS, a publically available dataset containing performances and their written critiques [4]. While this dataset does not provide real-time performance instruction, it is applicable to our study as it focuses on the instructor's "words."

2 Data Preparation

Features of Audio Segment of Classical Guitar Lessons In our previous study, we collected audio data from a pair (teacher and student) of one-on-one guitar lessons and categorized them into groups based on musical pieces for which the lessons had been given at least three times [1]. We then manually transcribed the utterances for each teacher and student and segmented them on the basis of the time interval between utterances.

Our analysis of the segmented data showed that (i) the percentage of students' utterances tended to decrease as the number of lessons progressed, and (ii) the percentage of teachers' utterances did not change much. These findings suggest that a more detailed analysis focusing on teachers' utterances might make it possible to extract the main points of the lesson. In addition, an integrated analysis of audio segments and the content of utterances is necessary to achieve a comprehensive summary of the lesson.

Annotation of Types of Teacher's Utterance In this study, we focused on the teacher's utterances and annotated two types of labels representing instructional information with the transcribed data. Through these annotations, it is possible to clarify the type of instructions given at specific times during a musical lesson.

– Instructional Topic Labels: The Music Teacher's Ontology [5] is a knowledge system for music education that represents a hierarchy of topics on which teachers might provide feedback to students. We referred to this ontology to define instructional topic labels. First, we defined four upper-level labels: *Musical piece*, regarding the musical style of the piece (period, musical form, etc.), *Musical expression*, regarding the intentional use of expression by the teacher or learner, *Technique*, regarding physical technique, and *Other*, regarding performance, mental aspects and so on. Then, we defined 18 specific topics as lower-level labels, such as *Tempo*, *Rhythm*, *Fingering*, and *Articulation*. We annotated one upper-level label and two or fewer lower-level labels for each unit of utterance.

– Instructional Content Labels: SOAP is a classification framework that takes into account the semantic elements of sentences for natural language [6]. It was originally designed for scientifically describing the thought processes of doctors in medical records using natural sentences. In this study, by referring to research [4], we adapted the classification categories of SOAP to the field of music as follows:

- Subjective data (*S*): teacher providing general and/or specific conceptual information based on subjectivity.

– Objective data (*O*): teacher providing general and/or specific conceptual information based on objectively referable events or concepts.
– Assessment (*A*): teacher's evaluation of a student's applied and/or conceptual knowledge.
– Plan (*P*): giving a specific opinion or recommendation to guide the student's action towards the achievement of a specific musical aims.

We pointed that the four items can characterize the instructional content and annotated them to the teacher's utterances. If an utterance had more than one label, we provided two or more annotations.

3 Structural Analysis of Teacher's Utterances

Generation of Tree Structures using the GTTM The Generative Theory of Tonal Music (GTTM) is a musical framework that can generate a tree structure with a hierarchical temporal organization and a principal-subordinate relationship of branches. A key feature of GTTM is that it enables somewhat subjective musical analysis because it allows for different interpretations based on the analyst's judgment while still adhering to the basic rules. It also enables "reduction," which extracts abstracted groups from the upper layers of the tree structure. These mean that GTTM can be utilized to reveal the analyst's perspective and identify structuring rules for summarizing the lesson.

In this study, two researchers independently generated a tree structure. First, we took one of one lesson data described in previous section and divided the teacher's utterances into five sections based on the segments and the content of the utterances. Next, we applied the time-span analysis method defined in the GTTM for grouping and hierarchization.

Comparison of Tree Structures and Discussion Figure 1 shows an example of utterances, annotated labels, and two generated tree structures, where structures A and B correspond to the results of analysts A and B. As we can see, the groupings of the two tree structures tended to be similar. Specifically, the branches were divided into three groups, suggesting a commonality in the analysts' perceptions of speech cohesion. On the other hand, there were differences in the hierarchy within the groups. This was due to the fact that (i) analyst A was conscious of summarization and placed *S* in the upper-level labels of the instructional content labels, representing the problems of the student's performance, whereas (ii) analyst B focused on the number of low-level labels and the latter half of the utterance.

The breakdown of instructional content labels in the five sections was *S*: 16.7%, *O*: 18.6%, *A*: 2.1%, *P*: 37.5%, *S&A*: 2.1%, and None: 22.9%. The fact that about 23% of the utterances could not be annotated (None) indicates that semantically significant utterances were limited. Examples of such unannotated utterances include rhythmic counting and responses using short words phrases as "I see." Furthermore, *P* tended to be located at a higher level of hierarchy in both structures. These results indicate that instructional content labels are useful for identifying hierarchies.

We need to collect additional tree structure data because the samples in this study were very small. Additionally, we need to define aggregation rules for the tree structures by applying GTTM: specifically, (1) Grouping Preference Rules, which group

Speaker	Utterances	Instructional topic label			Instructional content label
		Upper-level	Low-level	Low-level	
Teacher	Yes, about this much. One-two-three.	*Musical expression*	*Tempo*		
Teacher	Well, you tend to rush with the portamento.	*Musical expression*	*Tempo*	*Rendition*	*S*
Teacher	Yes.	*Musical expression*	*Tempo*	*Rendition*	
Teacher	Please keep the sound properly to the note value. This too.	*Musical expression*	*Rendition*	*Note value*	*P*
Student	Oh, I see.				
Teacher	Also, it is fast after "ti."	*Musical expression*	*Tempo*		*S*
Teacher	Keep the note value tight.	*Musical expression*	*Note value*		*P*
Teacher	Yes, I think that would make it in time. You are rushing too much about this.	*Musical expression*	*Tempo*		*S・A*
Student	Yes, around here, I'm in a hurry, so my performance is getting squished, here.				
Teacher	Flop.	*Musical expression*	*Tempo*		
Teacher	Yes, play each note properly.	*Musical expression*	*Tempo*	*Note value*	*P*
Student	I see. My performance is getting more and more messed up here.				
Teacher	Yeah, don't rush. As you get better at playing it, the speed gets faster. So, you have to be careful.	*Other*	*Tempo*		*A*

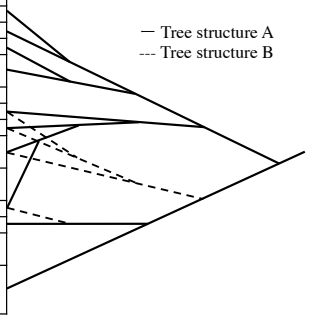

Fig. 1. Example of utterance data and tree structures.

utterances based on measures such as utterance duration, interval, and the number of utterances, and (2) Significance Preference Rules, which identify important utterances based on key words and label information.

4 Conclusion

In this paper, we performed a structural analysis of teacher's utterances to summarize the content of musical instrument lessons. First, we annotated the transcribed data of a classical guitar one-on-one lesson with two types of semantic labels. Then, we hierarchically structured the utterances based on the GTTM. As a result, we were able to clarify the analyst's perspective and examine the issues with the structuring rules. In future work, we will expand the data and conduct a more in-depth analysis of individual data. This will enable us to apply the findings to support student reflection and improve the quality of teaching.

References

1. Miura, H., Iino, N., Hamanaka, M., Takeda, H., and Nishimura, T.: An Attempt on Modeling of Teaching Knowledge in Musical Instrument Performance Situations, JSAI2021 (2021).
2. Lerdahl, F. and Jackendoff, R.: A Generative Theory of Tonal Music, MIT Press (1983).
3. Miura, H., Takegawa, Y., Terai, A., and Hirata, K.: Discussion Summarization Based on Hierarchical Structure Using Verbal and Non-Verbal Information, in Proceedings of the International Conference on Internet and Multimedia Technologies 2018, World Congress on Engineering and Computer Science, pp. 308–313 (2018).
4. Matsubara, M., Kagawa, R., Hirano, T., and Tsuji, I.: CROCUS: Dataset of Musical Performance Critiques Relationship between Critique Content and Its Utility, Proc. of the 15th International Symposium on CMMR, pp. 279–288 (2021).
5. Yee-King, M. J., Wilmering, T., Llano, M. T., Krivenski, M., and d'Inverno, M.: Technology Enhanced Learning: The Role of Ontologies for Feedback in Music Performance, Frontiers in Digital Humanities, **5**(29) (2019).
6. Weed, L. L.: Medical records, medical education, and patient care: The problem-oriented record as a basic tool, Press of Case Western Reserve University (1969).