A SCORE-INFORMED PIANO TUTORING SYSTEM WITH MISTAKE DETECTION AND SCORE SIMPLIFICATION

Tsubasa Fukuda Yukara Ikemiya Katsutoshi Itoyama Kazuyoshi Yoshii
Graduate School of Informatics, Kyoto University
{tfukuda, ikemiya, itoyama, yoshii}@kuis.kyoto-u.ac.jp

ABSTRACT

This paper presents a novel piano tutoring system that encourages a user to practice playing a piano by simplifying difficult parts of a musical score according to the playing skill of the user. To identify the difficult parts to be simplified, the system is capable of accurately detecting mistakes of a user's performance by referring to the musical score. More specifically, the audio recording of the user's performance is transcribed by using supervised non-negative matrix factorization (NMF) whose basis spectra are trained from isolated sounds of the same piano in advance. Then the audio recording is synchronized with the musical score using dynamic time warping (DTW). The user's mistakes are then detected by comparing those two kinds of data. Finally, the detected parts are simplified according to three kinds of rules: removing some musical notes from a complicated chord, thinning out some notes from a fast passage, and removing octave jumps. The experimental results showed that the first rule can simplify musical scores naturally. The second rule, however, often simplified the scores awkwardly when the passage formed a melody line.

1. INTRODUCTION

Thanks to the recent development of audio signal analysis technology, many applications have appeared that enable users to practice playing musical instruments without the guidance of a teacher. A system called SongPrompter [1], for example, automatically displays the information (*e.g.*, chord progression, tempo, lyrics) for assisting a user to play a guitar and sing a song. An application [2] estimates the chord progressions of user's favorite songs taken from an iPhone or iPod and creates chord scores for them.

In this paper we propose a novel piano tutoring system that can detect mistakes of piano performances and simplify the difficult parts of musical scores[1] because the piano is one of the most popular musical instruments. Although players at an intermediate level want to play their favorite musical pieces, the scores of those pieces are often difficult for those players to play, causing them to lose their motivations. One of the effective solutions for this problem is to simplify the musical scores so that the difficulty of those scores matches the user's playing skills. To effectively assist a user to improve his or her playing skill, it is important to gradually increase the difficulty level of a musical score to recover the original difficulty level by changing the score simplification level. As the first step toward this goal, in this paper we focus on how to simplify mistakenly-played parts of musical scores.

The proposed system takes an audio signal of users' actual performance and the original score as inputs, and outputs a piano roll that shows user's mistakes and a simplified version of the original score. First, two piano rolls are created from the input audio signal and musical score respectively. More specifically, one is converted from the audio signal by using a multipitch estimation method based on nonnegative matrix factorization (NMF), and the other is obtained by synchronizing the original score with the users' performance with dynamic time warping (DTW). User's mistakes are then detected by comparing these two piano rolls, and the original score is simplified in accordance with the parts in which the mistakes are detected. We define three kinds of rules for simplifying musical scores and how to apply those rules.

Two experiments using actual music performances were conducted to evaluate the performance of the proposed system. The first experiment focused on the accuracy of mistake detection. Although double or half octave errors are the main cause of decreasing accuracy in multipitch estimation, those errors can be ignored for the purpose of detecting performance mistakes because it is rare for a user to mistakenly play double- or half-pitch notes. The second experiment focused on the effectiveness of score simplification. The results showed that some musical notes can be removed naturally from complicated chords and that removing musical notes from fast passages should be avoided when the passage constituted a melody line.

2. RELATED WORK

This section introduces related work on multipitch estimation and score simplification.

2.1 Multipitch estimation and mistake detection

It is necessary for revealing a user's weak points to detect mistakes by comparing the result of multipitch estimation and the original score.

[1] A demo video is available on http://winnie.kuis.kyoto-u.ac.jp/members/tfukuda/smc2015/

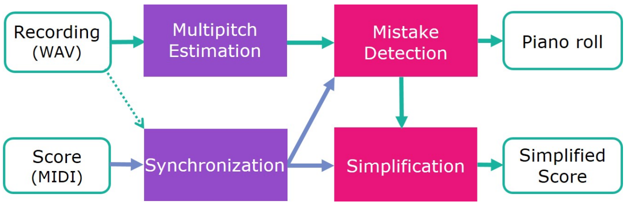

Figure 1. An overview of the proposed system

Tsuchiya *et al.* [3] proposed a novel Bayesian model that combines acoustic and language models for automatic music transcription. They tested the model on the RWC music database [4] and showed the result of transcription. They categorized transcription errors into three types, that is, deletion errors, pitch errors, and octave errors. As shown in the result of an experiment, octave errors are the majority of detected errors.

Azuma *et al.* [5] proposed a method of automatic transcription for a piano performance with both hands by focusing on harmonic structures in the frequency domain. This method automatically separates an obtained score into melody and accompaniment parts by focusing on the probability of the pitch transition. This system takes only an audio signal as an input, and many octave errors occur in the result of transcription.

Emiya *et al.* [6] and Sakaue *et al.* [7] showed that modeling the harmonic structures of musical instruments improves the accuracy of multipitch estimation. Since piano sounds also have the harmonic structures, the accuracy improvement is expected by integrating the prior information of harmonic structures into our system.

Benetos *et al.* [8] proposed a score-informed transcription method for automatic piano tutoring. Although the F-measure of automatic transcription was about 95%, many octave errors occurred. Our main contributions is to take into account those errors for mistake detection and to combine mistake detection [8] with score simplification for effectively assisting a user to practice playing the piano.

2.2 Score simplification

Simplifying a musical score according to the player's skill motivates him or her to practice the piano effectively. Just removing the notes from the score is, however, insufficient, for the score simplification. Since it is necessary to preserve the characteristics of the original score, how to simplify the score is a very important problem.

Yazawa *et al.* [9] proposed a method of guitar tablature transcription from audio signals. In this method, playing difficulty costs are given to several features such as positions of player's hands, the number of fingers used, and the migration length of the wrist. Then, it creates a tablature matching player's skills based on the cost.

Fujita *et al.* [10] proposed a method that modifies a musical score consisting of several instrument parts according to the player's skill. Player's skills are categorized into three types, and simplification is done based on four factors *i.e.*, the number of simultaneous notes, the width of a chord, the width of a passage, and the tempo of a passage.

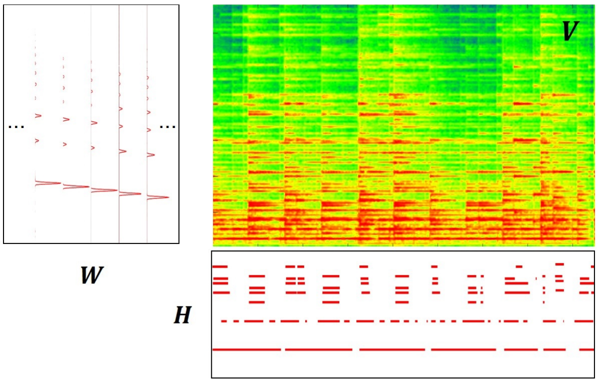

Figure 2. Multipitch estimation using NMF. An activation matrix H is calculated from an input spectrogram V by using a pretrained basis matrix W.

3. PROPOSED SYSTEM

This section describes a proposed system that can simplify difficult parts of a musical score according to the playing skill of the user. An overview of the proposed system is shown in Figure 1. The inputs are 1) an audio recording of a user's piano performance and 2) the original musical score. The outputs are 1) a piano roll indicating mistakes and 2) a simplified score.

The score is simplified by first calculating an activation matrix from the input recording using non-negative matrix factorization (NMF). The activation matrix is then converted into a piano roll by thresholding. The musical score is next synchronized with the audio recording by stretching the onset times and duration of the musical notes using dynamic time warping (DTW). The synchronized score is then converted into a piano roll.

Mistakes are detected by comparing the two synchronized piano rolls. Detection accuracy is improved by ignoring the octave errors that rarely occur during actual playing. The parts where mistakes were detected are classified into three patterns and simplified in accordance with predefined simplification rules.

3.1 Multipitch estimation

The user's playing performance is evaluated using the result of multipitch estimation. The input is the recording, and the output is a piano roll of the recording.

The estimation is done using the NMF algorithm with β-divergence [11]. This algorithm factorize a matrix $\boldsymbol{V}$ is factorized into two matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ ($\boldsymbol{V} = \boldsymbol{WH}$) that have no negative elements.

First, the recording is converted into a spectrogram as a matrix $\boldsymbol{V} \in \mathbb{R}^{f \times n}$ using constant-Q transform (CQT) [12], which has 24 frequency bins per octave and can handle frequencies as low as 60 Hz. The NMF algorithm takes matrix $\boldsymbol{V}$ as input and factorizes it into $\boldsymbol{W}$ and $\boldsymbol{H}$. Here, $\boldsymbol{W} \in \mathbb{R}^{f \times 88}$ is the base spectrum matrix. It consists of 88 base spectra from A0 to C8. The activation matrix is $\boldsymbol{H} \in \mathbb{R}^{88 \times n}$. It contains the amplitudes of the base spectra. NMF typically factorizes $\boldsymbol{V}$ by iteratively updating $\boldsymbol{W}$ and $\boldsymbol{H}$. Since $\boldsymbol{W}$ is fixed here, only $\boldsymbol{H}$ is updated in accordance with the following rule:

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \otimes \frac{\boldsymbol{W}^T((\boldsymbol{V} \otimes \boldsymbol{WH})^{\beta-2})}{\boldsymbol{W}^T(\boldsymbol{WH})^{\beta-1})}, \qquad (1)$$

where $\otimes$ is the element-wise product, the exponentiation is element-wise exponential and the fraction means element-wise division. We used $\beta = 0.6$, which has been shown to produce the best multipitch estimation of piano sounds in previous studies [8, 11, 13].

Base spectrum matrix $\boldsymbol{W}$ is estimated in advance from the sound of each pitch using an electronic piano. Application of CQT to each recording produces 88 spectrograms $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{88}$. NMF with a single basis spectrum is applied to each $\boldsymbol{X}_i$, *i.e.*, $\boldsymbol{X}_i \in \mathbb{R}^{f \times n}$ is factorized into vectors $\boldsymbol{w}_i \in \mathbb{R}^{f \times 1}$ and $\boldsymbol{h}_i \in \mathbb{R}^{1 \times n}$ as follows:

$$\boldsymbol{X}_i = \boldsymbol{w}_i \boldsymbol{h}_i. \qquad (2)$$

Base spectrum matrix $\boldsymbol{W}$ is finally obtained by concatenating these vectors horizontally as follows:

$$\boldsymbol{W} = [\boldsymbol{w}_1 \boldsymbol{w}_2 \cdots \boldsymbol{w}_{88}]. \qquad (3)$$

After update rule (1) converges, a piano roll of the recording is obtained by thresholding activation matrix $\boldsymbol{H}$ appropriately. An example of the NMF results is shown in Figure 2. The sample musical piece is from the RWC music database.

3.2 Score-to-audio synchronization

We obtained the audio recording using a YAMAHA P-255 electronic piano which can record an actual performance as an audio signal. There were several temporal gaps between the musical score and the actual performance no matter how exactly it was played. If such gaps are counted as mistakes, true mistakes cannot be detected appropriately. We avoid this problem by synchronizing the musical score with the recording in advance. The inputs are 1) the recording and 2) the musical score, and the outputs are the corresponding synchronized piano rolls.

For synchronization, we use the dynamic time warping (DTW) algorithm of Muller [14] to measure the similarity between two temporal sequences. Since this algorithm is a kind of dynamic programming, it can obtain the optimum solution, which means that the temporal correspondence between these sequences can be obtained by using it. Specifically, inputs are converted into spectrograms in advance. These spectrograms are next converted into chroma vectors, which are the 12-dimensional vectors. A chroma vector has the amplitude of each pitch name (C, C#,. . .,B), and the cosine distance of two chroma vectors are used as the distance in DTW.

3.3 Mistake detection

Comparing the piano roll created from the recording with the synchronized piano roll reveals where the user played the song incorrectly. The inputs are 1) the piano roll from the recording and 2) the synchronized piano roll, and the output is a piano roll comparing the inputs. The system compares the two input scores and indicates where the user made mistakes in his or her performance. An example output piano roll is shown in Figure 3. The black marks correspond to the notes the user played, and red marks correspond to notes in the original score. There were two mistakes in this example, as shown by the two blue marks.

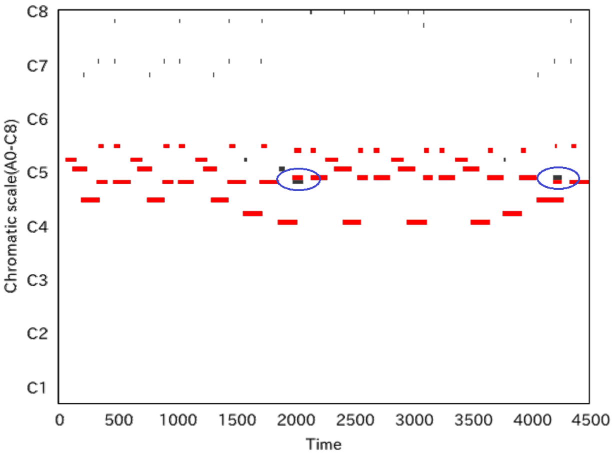

Figure 3. Example of mistake detection. Red marks correspond to the notes in an original score and black marks correspond to the notes estimated by NMF.

The system judges the weak points in the user's performance on the basis of where the mistakes are in the recording. Since octave errors often occur in multipitch estimation using NMF, the system sometimes misjudges the location of the mistakes. In fact, many short notes (around C7 and C8) that were not actually played by the user are detected in Figure 3. Since playing these notes rarely occur in the actual performance of a piano solo, we ignore octave errors to improve the accuracy of the multipitch estimation. Specifically, a detected mistake is ignored if there is another note whose pitch differs from the pitch of the detected mistake by octaves.

3.4 Score simplification

Simplifying the difficult scores to match the user's playing skills helps motivate the user to practice the piano. The inputs are 1) the original score and 2) the parts to be simplified, and the output is the simplified score.

In this part, scores are classified under three patterns, and are simplified according to simplification rules given in advance for each pattern.

Pattern 1. Parts with many notes to be played at the same time

Pattern 2. Parts that require fast fingering

Pattern 3. Parts that have adjacent notes, one is over an octave distant from the other.

Here we describe simplification process by using examples.

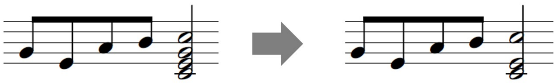

Figure 4. Example of pattern 1. Removing some notes from chords.

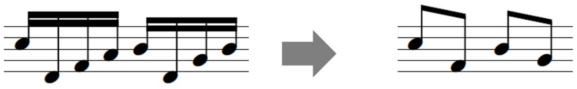

Figure 5. Example of pattern 2. Removing some notes from a part that requires fast fingering.

3.4.1 Pattern 1: Simplifying chords

When there are many notes to be played at the same time, that part is simplified removing some notes from the chords, as shown in Figure 4. Priority is given to each pitch of the chord, and the notes with lower priority are removed.

More specifically, the melody line often consists of a note with the highest pitch of the chord and is one of the most important notes. A note with the lowest pitch of the chord, called the root of the chord, is also important. These two notes are especially important and thus should not be removed, as shown in previous study [15]. The other notes are less important and can be removed if necessary. As a result, chords that are difficult to play are simplified.

3.4.2 Pattern 2: Simplifying fast passages

A part that requires fast fingering is simplified by removing the sequential notes that are to be played faster than a threshold, as shown in Figure 5.

3.4.3 Pattern 3: Removing octave jumps

Parts that have adjacent notes, one is over an octave distant from the other are called "leaps" and are further classified into two patterns.

3-A There is another leap after the leap.

3-B There is no leap after the leap.

In pattern 3-A, a note that generates the leap is difficult to play, so it would be removed as shown in Figure 6 (a).

In pattern 3-B, notes around the leap are removed in accordance with the rule for pattern 1 or 2. That is, if there is a chord around the leap, it is simplified, and if there is a fast passage around the leap, it is simplified as shown in Figure 6 (b).

4. EVALUATION

This section reports two experiments that were conducted for evaluating the performance of score-informed multipitch estimation and score simplification.

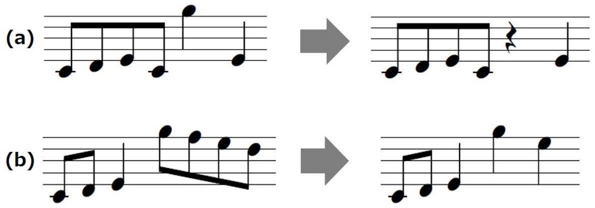

Figure 6. Examples of pattern 3. Removing some notes around a leaping.

Octave errors	Precision	Recall	F-measure
NOT ignore	0.943	0.988	0.965
ignore	0.995	0.988	0.991

Table 1. Accuracy of multipitch estimation

4.1 Multipitch estimation

We calculated the accuracy of multipitch estimation by comparing two piano rolls. One was obtained by analyzing the recording of an actual performance, and the other was created from an original musical score. The audio signals were recorded using a YAMAHA P-255 electronic piano played by a intermediate player, and were converted into a spectrogram using CQT which had 24 frequency bins per octave. This spectrogram was then factorized into a basis spectrum matrix and an activation matrix by using NMF. The activation matrix was finally converted into a piano roll by thresholding. On the other hand, pitch, onset time, and duration of each note were obtained by the original score, and the correct piano roll was created by synchronizing with the actual performance using DTW.

As shown in Table 1, the F-measure was calculated by comparing those piano rolls while ignoring octave errors and in the not case by way of comparison. About first ten seconds of *The Flea Waltz* was used as test data. According to the result, the F-measure was improved by ignoring octave errors. Using an appropriate value in thresholding helps to obtain the high accuracy.

We plan to employ a more reliable method for binarization of an activation matrix that is obtained by NMF. More specifically, a hidden Markov model (HMM) can be employed instead of thresholding. This model helps to reduce very short notes that are often occurred as octave errors in the result of NMF algorithm.

4.2 Score simplification

We evaluated the effectiveness of score simplification. The Grande valse brillante in E-flat major, Op. 18 and Étude Op. 10, No. 12 were used for this evaluation. Here, we tried simplifying parts that were selected at random on the assumption that the parts were played incorrectly by a user.

As simplifying a score, we prepared the score data that has pitch, onset time, and duration of each note. This time we chose the last twenty seconds of the sample music and simplified them. Simplified scores are shown in Figure 7

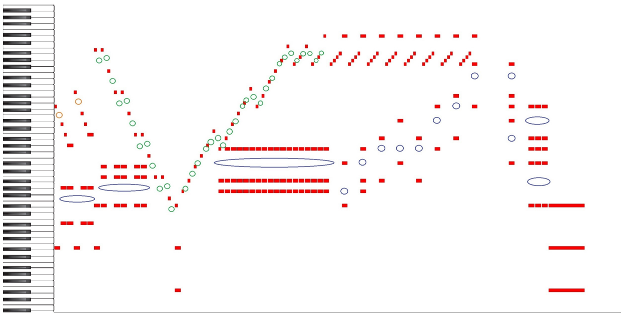

Figure 7. Simplified score of The Grande valse brillante in E-flat major, Op. 18. Blue marks correspond to the notes simplified in pattern 1, green ones correspond to pattern 2, and orange ones correspond to pattern 3.

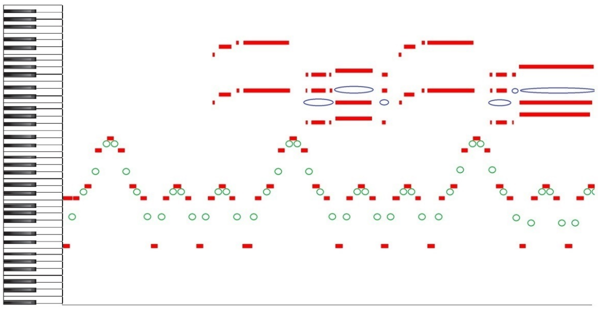

Figure 8. Simplified score of Étude Op. 10, No. 12

and Figure 8. In these figures, blue marks correspond to the notes simplified in pattern 1, green ones correspond to pattern 2, and orange ones correspond to pattern 3.

According to the result, we found that the simplification was correctly done. It was felt that something was a little off about simplification in pattern 2 and there was room for improvement. In patterns 1 and 3, it was felt that the result of simplification was naturally done. We will focus on appropriateness for the rules of simplification and automation of them in future work.

5. CONCLUSION

Our score-informed piano tutoring system with mistake detection and score synchronization works well by analyzing a recording of an actual performance. Intermediate players are often faced by a problem that the scores of their favorite pieces are often difficult to play and this makes them lose their motivations. The proposed system helps those players effectively continue to practice playing the piano. It detects the parts of the score that should be simplified so that the user can easily play those parts. Since the system can use the audio recording of a user's performance, it is unnecessary to set detailed parameters. Since the kinds of scores in which playing errors are likely to be occurred are identified to a certain level, the proposed system categorizes those errors into three types and simplifies the score according to predefined rules for each type.

Possible future works on the proposed system are as follows:

- Carry on additional experiments
- Improve each algorithm
- Improve score simplification

First, the amount of experiments is insufficient and there is a possibility that the evaluation is incorrect. We have to carry on additional experiments for various conditions to confirm that the evaluation is appropriate. It is also necessary to carry on an experiment through the whole system.

Second, improving each algorithm is necessary. DTW, employed in the synchronization, obtains optimal solution, but is a bit computationally expensive. An alternative solution is to use windowed time warping (WTW) [16]. Although this algorithm requires the distant paths to be contained in the correct paths, this requirement is met between an actual performance and the original score.

Finally, score simplification could be improved by using a wide variety of criteria. Future work will focus on the fingering to detect the notes that are difficult to play.

Acknowledgments

This study was supported in part by the OngaCREST project and JSPS KAKENHI 24220006, 26700020, and 24700168.

6. REFERENCES

[1] M. Matthias, H. Fujihara, and M. Goto, "SongPrompter: An accompaniment system based on automatic alignment of lyrics and chords," in *ISMIR2010*, 2010.

[2] CASIO COMPUTER CO., LTD., "Chordana viewer," http://world.casio.com/emi/app/ja/viewer/, 2015.

[3] H. Kameoka, K. Ochiai, M. Nakano, M. Tsuchiya, and S. Sagayama, "Context-free 2D tree structure model of musical notes for Bayesian modeling of polyphonic spectrograms." in *ISMIR2012*, 2012, pp. 307–312.

[4] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Popular, classical and jazz music databases." in *ISMIR2002*, vol. 2, 2002, pp. 287–288.

[5] M. Azuma and W. Mitsuhashi, "Automated transcription for polyphonic piano music with a focus on harmonics in log-frequency domain," *IPSJ SIG Technical Reports*, vol. 89, no. 28, pp. 1–6, 2011.

[6] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 18, no. 6, pp. 1643–1654, 2010.

[7] D. Sakaue, K. Itoyama, T. Ogata, and H. G. Okuno, "Initialization-robust multipitch estimation based on latent harmonic allocation using overtone corpus," in *ICASSP2012*, 2012, pp. 425–428.

[8] E. Benetos, A. Klapuri, and S. Dixon, "Score-informed transcription for automatic piano tutoring," in *EUSIPCO 2012*, 2012, pp. 2153–2157.

[9] K. Yazawa, D. Sakaue, K. Nagira, K. Itoyama, and H. G. Okuno, "Audio-based guitar tablature transcription using multipitch analysis and playability constraints," in *ICASSP2013*. IEEE, 2013, pp. 196–200.

[10] K. Fujita, H. Oono, and H. Inazumi, "A proposal for piano score generation that considers proficiency from multiple part," *IPSJ SIG Technical Reports*, vol. 77, no. 89, pp. 47–52, 2008.

[11] A. Dessein, A. Cont, and G. Lemaitre, "Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence," in *ISMIR2010*, 2010, pp. 489–494.

[12] C. Schörkhuber and A. Klapuri, "Constant-Q transform toolbox for music processing," in *SMC 2010*, 2010, pp. 3–64.

[13] J. Fritsch and M. Plumbley, "Score informed audio source separation using constrained nonnegative matrix factorization and score synthesis," in *ICASSP2013*, 2013, pp. 888–891.

[14] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*. Springer, 2007, pp. 69–84.

[15] G. Hori, H. Kameoka, and S. Sagayama, "Input-output HMM applied to automatic arrangement for guitars," *Information and Media Technologies*, vol. 8, no. 2, pp. 477–484, 2013.

[16] R. Macrae and S. Dixon, "Accurate real-time windowed time warping." in *ISMIR 2010*, 2010, pp. 423–428.