Automatic Electronic Organ Reduction Using Melody Clustering

Daiki Tanaka

Graduate School of Computer and Informaion Sciences
Hosei University
Tokyo, Japan
Email: 17t0016@cis.k.hosei.ac.jp

Katunobu Itou

Faculty of Computer and Information Sciences
Hosei University
Tokyo, Japan
Email: itou@hosei.ac.jp

***Abstract*—Reduction is a method for arranging the scores of a multipart composition to its ensemble. In this study, we propose an arrangement system using the full score to automate the arrangement by reduction. Our target instrument for this study is an electronic organ, which has a score similar to the score of an ensemble. First, we performed clustering based on the rhythm, melodic activity, harmony, sonic richness, and timbre of the instruments to reduce the part number of the melody in the full score. Further, we selected clusters of melodies corresponding to the right-hand, left-hand, and foot parts of the electronic organ. Finally, the musical score was rectified to be able to play the electronic organ. This system was evaluated using four songs. The average values of the right-hand, left-hand, and foot parts were 0.80, 0.71, and 0.77, respectively. These results depicted that the proposed arrangement was suitable for preparing electronic organ scores from the full score.**

Keywords–Music; Arrangement; Reduction; Clustering; Electronic Organ.

I. INTRODUCTION

Piano and electronic organs can play orchestral songs and can be used to express the orchestra performance. However, all songs do not have musical scores. Therefore, we proposed a system that can automatically generate music scores for different songs.

Several studies have been performed on the piano music arrangement; however, the electronic organ music arrangement has not been much studied. Electronic organ can be used to play a wide range of sounds and rhythms. Nonetheless, the number of electronic organ notation is less than the number of piano score. Therefore, this study aims to focus on electronic organ music arrangement using an arrangement method known as reduction. In reduction, the full score (shown in Figure 1) composed of multiple parts is contracted to the main elements by reducing or eliminating the melody. When arranging by reduction, the system must consider the melodies that it must eliminate. These melodies change depending on the instruments for which the music is being arranged; therefore, it is difficult to definitely set any criteria. For example, the piano score comprises two parts: the right and left hands. To reduce a piano score, we can omit a melody like the main melody and the accompaniment. For instruments like the guitar, we eliminate the chord melody. In this way, reduction is performed based on the characteristics of the musical instrument whose scores have to be arranged.

The remainder of the paper is structured as follows. Section II introduces the characteristics of electronic organ and explains the related works. Section III introduces the arrangement method of electronic organ score. Section IV gives the test results of arrangement system. Section V discusses the results. Conclusions are given in Section VI.

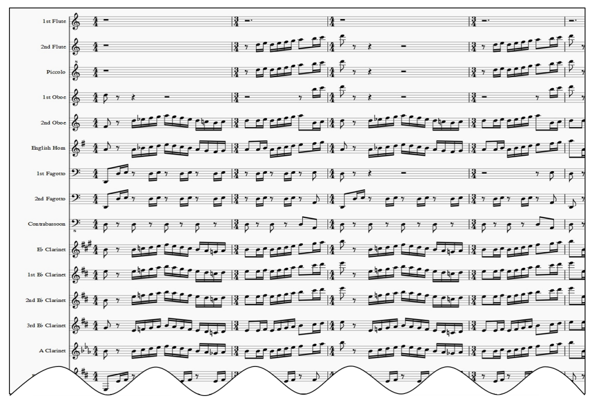


Figure 1. A Multipart Musical Score

II. ELECTRONIC ORGAN REDUCTION

A. Characteristics of the electronic organ

The electronic organ is played using the right and left hands, whereas an electronic organ is played using the right hand, the left hand, and a foot. Additionally, the piano has one keyboard, but the electronic organ has three keyboards. Therefore, the score for an electronic organ is a three-set score (e.g., see Figure 2).

The right hand of the melody often constitutes the main melody. The left hand of the melody often contains the chords with sound numbers of four or less. Foot melodies mainly comprise the root of the chord and are always single notes. The pitch is observed to initially reduce for the right-hand melodies, then for the left-hand melodies, and finally for the foot melodies. When the right-hand and left-hand melodies coincide in time, their range is within one octave.

As a major feature of an electronic organ, it is possible to reproduce timbres for approximately 900 types of musical instruments. Therefore, different timbres overlap on a single keyboard. Additionally, it is possible to switch timbres automatically; therefore, we can continue playing the melodies of different instruments.

Electronic organs contain rhythm boxes, such as percussion instruments. The music rhythms are prepared in advance, and the player plays the organ using this rhythm.

Figure 2. Electronic Organ Score

B. Related Work

Using reduction, the scores of a multipart music can be condensed to the musical score of a single target instrument. Several studies have been performed on reduction by automatic arrangement. Fujita [1] proposed a method for creating piano notations using multiple parts. The method focuses on the average pitch, pronunciation time, pitch, and rhythm pattern and estimates the melody and baseline from the melody of the full score. Additionally, we adopted the melodies for the right-hand and the left-hand parts of the piano. However, we were unable to uniquely determine the melody for the polyphonic songs using this method. Furthermore, the piano does not always continue the melody and baseline at all times. Therefore, this method is insufficient. In this research, we decided to put together the melodies with similar characteristics instead of estimating and scoring the melody. Using these melodies, we chose the melody that sounded similar to that of a piano.

Ito et al. [2] intended to make an ensemble music notation from the full score. First, they performed the clustering of melodies. Further, they eliminated unnecessary melody clusters. However, this clustering did not consider the musical point of view because the researchers considered only the distance of the notes on the score. Matsubara et al. [3] studied clustering in consideration of the musical features. They performed clustering using the features of rhythm, melodic activity, harmony, and sonic richness. In this study, we added a new feature, the timbre quantity, to the clustering proposed by Matsubara et al. Further, using the clustered melodies, we selected a melody cluster that reflected the characteristics of the electronic organ and created an electronic organ score.

Rose Curtis [4] stated that the timbre of an instrument depends on the amplitude envelope of the sound, fluctuations caused by vibrato and tremolo, formant structure, perceived volume, duration, and temporal frequency of the component fluctuation.

Therefore, we decided to focus on the formant structure. The common methods used to analyze the formant structure are the Linear Predictive Coding (LPC) [5], LPC cepstrum [6], cepstrum, and Mel Frequency Cepstrum Coefficient (MFCC) [7]. In this study, we analyzed the sounds of different musical instruments and the sound range that could be produced by each instrument. It was difficult to compare within the same note range. In addition, the sound produced by the musical instrument has the property of a harmonic structure in which the frequency of the integral multiples of the fundamental frequency appears strongly. Thus, the LPC for obtaining the formant structure including the portion where the frequency strongly appears is not suitable here. In this study, we analyze the timbre of musical instruments using MFCC as an analysis method to investigate the formant structure that is not considerably affected by the pitch.

III. ARRANGEMENT METHOD OF ELECTRONIC ORGAN SCORE

This system inputs the full score expressed in the MusicXML [8] format. The system must obtain the instrument name, bar number, octave, note type,pitch, duration, and metrical information from the MusicXML file. Based on this information, we prepared the five features (the pronunciation time, pitch change, harmony, persistent pronunciation pattern, and timbre of the instrument). The full score of the melody was clustered using these features and was further grouped into multiple melody groups having similar characteristics. After clustering, we selected the three groups corresponding to the right hand, left hand, and foot parts of the electronic organ score. Finally, the score was corrected so that it could be played on an electronic organ. The electronic organ score arranged by the system generated the musical score having three parts and was written out in the MusicXML format. We used the MuseScore [9] scorewriter software for creating the MusicXML format.

A. Melody Grouping

1) Features: The melodies are classified based on five features: rhythmic activity, melodic activity, consonance activity, sonic richness, and instrument timbre.

Based on the literature [3], we prepared four features: rhythmic activity, melodic activity, consonance activity, and sonic richness. The timbre feature of the instrument is a newly defined quantity in this research.

Rhythmic Activity: This feature considers the amount of rhythm in the musical score. When the two phrases in a melody are pronounced at the same time, they possess the same rhythm. For example, the rests after the beats in Figure 3(b) generate the same rhythm as depicted in Figure 3(a). As the basic unit, we used the shortest note per unit time. This unit was assigned the value 1 when the note was sounded, and the value 0 was assigned when the note was not sounded. This generated a binary vector RA with n elements.

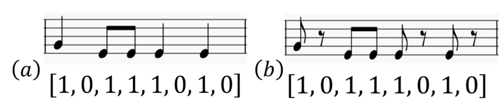


Figure 3. Feature vector of rhythmic activity

Melodic Activity: For most melodies, the pitch varies throughout the melody. Therefore, as the harmony often shares the same movements as the melody, the harmonies and melodies are often considered to be the same phrase. For example, the melodies in Figure 4(a) and (b) can be regarded to be similar melodies with different pitches. In each unit time, a transition between two notes is set to 1 when the note is higher than the preceding note. The transition is set to 0 when the pitch does not change and to -1 when the note is lower than the preceding note. This process generates a vector MA of melodic activity. The vector MA compares each note with its previous neighbor. There is no sound previous to the head sound. Therefore, the comparison starts from the note that immediately follows the head note. This means that the vector is $n - 1$, i.e., the number of elements minus one.

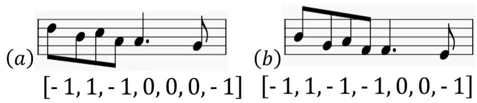


Figure 4. Feature vector of melodic activity

Consonance Activity: Even when the pitches follow similar movements, many non-harmonious sounds can appear in the parts of the melody. These harmony-like accompaniment parts are recognized as different phrases. Therefore, we calculated the top three sounds from all parts of the phrase per unit time and approximately replicated the harmony sound, as depicted in Figure 5(a). In each unit time, we set the inclusion or exclusion of harmonious tones to 1 and 0, respectively, generating a binary vector CA of consonance activity with n elements. Therefore, it is possible to distinguish the melody of Figure 5(c), which contains many melodies and harmonies, from the melody of Figure 5(b), which contains few harmonious sounds.

Figure 5. Feature vector of consonance activity

Sonic Richness: This feature quantifies the sound of the instrument. It is possible to distinguish the melody in Figure 6(b), which includes the melody and the rest elements;these elements can be easily distinguished from Figure 6(a) that does not include any rest elements. In each unit time, the presence or absence of sound is assigned as 1 and 0, respectively, generating a binary vector SR of sonic richness with n elements.

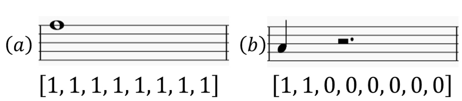


Figure 6. Feature vector of sonic richness

Instrument Timbre: Musical instruments of a similar timbre often play a similar melody. In an electronic organ, you can play with the timbres of multiple instruments stacked on top of each other. Therefore, it is common to play such melodies on the same keyboard. To play a melody, the feature vector IT (Instrument Timbre) is used to measure the similarity between the timbres of each instrument.

First, we modeled the timbres. To extract the features of the timbre, we prepared Musical Instrument Digital Interface (MIDI) sound sources of 29 musical instruments. These instruments are usually used in orchestras and brass performances. The MIDI sounds that we have used were the sounds produced for a time period of approximately 3 s. In addition, the timbre of the instrument could be changed by the pitch using a pitch of approximately two octaves of the note range that were used in the full score.

Further, we considered the section used by the MIDI sound source. The timbre is observed to alter from the beginning to the end of the sound produced by the instrument. Therefore, we decided to use the section corresponding to the Attack-Decay-Sustain-Release (ADSR) envelope [4] separately. ADSR is defined by dividing the amplitude envelope of the time waveform into four sections. The time required for the sound to suddenly move from nil to a peak is known as the attack. Decay is the interval in which the sound decays from the attack level to the designated sustain level. Sustain is the level at which the sound continues for as long as the vocalist's breath continues or while the piano/organ key is pressed. Finally, the release time is the time that is required to decay from the sustain level to zero. In this study, we used 1.5 s as the sustain interval, which does not change much with timbre and is relatively stable as the feature quantity. The formant structure was obtained using a 15-dimensional MFCC having a frame length of 256 points and a shift length of 128 points. This is used as a feature quantity of the timbre model.

We modeled the timbre using the obtained feature quantity. For modeling, we used the mixed Gaussian model called Gaussian Mixture Model (GMM) [10]. We designed the GMM with a mixture ratio of 8.

Further, we measured the similarity of the timbre. The timbre is modeled on the sum of multiple multidimensional normal distributions using GMM. In this study, the similarity of the timbre model is measured by comparing the shape of the probability density function. To calculate the similarity, we used the Kullback-Leibler (KL) divergence [11] for measuring the difference in the probability distribution. When the probability distributions of the timbres of the instruments of parts i and j are P and Q, which are continuous probability distributions, Equation (1) expresses the similarity using the KL divergence. Then, $p(x)$ and $q(x)$ are the probability density functions of the timbre of each musical instrument. A property of the KL divergence is that it is always 0 or more, and it is 0 when the two distributions are equal. The KL divergence is asymmetric (Equation (2)); therefore, we use the average of the KL divergence values in this study.

$$D_{KL}(i,j) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \tag{1}$$

$$D_{KL}(P||Q) \neq D_{KL}(Q||P) \tag{2}$$

The feature quantity of the timbre is a pairwise value between musical instruments. However, the feature quantity defined in [3] is the feature quantity of each musical instrument. Therefore, it is necessary to convert the timbre feature quantity to the feature quantity of each musical instrument. In this study, the value of the KL divergence was converted into the value of each instrument using Multi Dimensional Scaling (MDS) [12]. We converted the divergence to 5-dimensional features. In this study, the IT vector is a normalized vector of these feature vectors.

Figure 7 depicts the IT vector using an isomap in a 2-dimensional musical instrument space; it was visualized to

 ISBN: 978-1-61208-627-9

make the timbre distances of each instrument easier to understand. For example, a brass instrument trumpet and cornet are similar instruments; therefore, they have similar timbres. Therefore, their timbre values are located close to each other. Other instruments having similar timbres are a cello and viola among stringed instruments and an alto and tenor sax among woodwind instruments. Thus, the model created using tone colors can be the correct model.

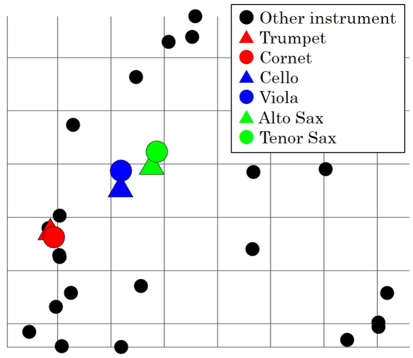


Figure 7. Visualization of the distance of timbre in musical instruments in instrument space

2) Similarity: The scale used for clustering is defined. For distance clustering, we adopted cosine similarity, which groups similar melodies. In the classification, the cosine similarity was subtracted from 1.

The total of all the weights $w_k(k = 1, 2, 3, 4, 5)$ is 1. $D_{RA}(i, j)$ defines the distance between the parts i and j using the feature vector RA. Similarly, $D_{SR}(i, j)$, $D_{MA}(i, j)$, $D_{CA}(i, j)$, and $D_{IT}(i, j)$ defines the distance using the feature vector SR, MA, CA, and IT. We performed clustering using the value of D given in Equation (3).

$$\begin{aligned} D(i,j) &= w_1 D_{RA}(i,j) + w_2 D_{SR}(i,j) \\ &+ w_3 D_{MA}(i,j) + w_4 D_{CA}(i,j) + w_5 D_{IT}(i,j) \end{aligned} \quad (3)$$

3) Melody clustering: We performed clustering by the k-means method using the above five features. The clustering was performed in units of two measures so that the number of simultaneous sounds remained below 5. Melody clustering was implemented as follows.

1) We set the cluster number to $K = 5$, and the part number as N. We acquired the note data for each part in two measures.
2) We prepared the feature vector defined in III-A1 from the note data of each extracted part and applied the k-means method with the cluster number K, using the scale defined by Equation (3).
3) We obtained the clustering results. The clustering terminated when there were four or fewer note names in each group simultaneously or when $K = N$; otherwise, we set $K = K + 1$ and returned to the first step.

The initial value was determined randomly. From the second time onward, the initial value was the value immediately preceding the label number. As the number of clusters differed for each measure, more clusters were assigned than the number of parts. After the experimental exploration, the initial number of clusters was set to 5.

B. Selection of classified groups

Among the classified groups, we identified the melodies corresponding to the right-hand, left-hand, and foot parts of the electronic organ score. We first selected the foot part (the most dissimilar playing method); this was followed by the right-hand and left-hand parts.

The main objective of the foot part was to supplement the whole song with bass melodies. Therefore, in each of the classified groups, melodies with the largest rate of notes with a pitch of C3 or less were selected for the foot part. The rates of bass melodies were calculated by dividing the number of notes below C3 by the number of notes per unit time.

Further, we selected the group corresponding to the right hand. The right hand produces higher sounds than the other keys; therefore, this selection considers the proportion of the treble notes. Furthermore, the right-hand part often controls the melody with a large number of pronunciations (the main melody); therefore, we also considered the rate of the number of pronunciations. The high pitch rate HR(High Rate) was calculated by dividing the number of notes above C5 by the number of notes per unit time. The sound-number ratio PR(Pronunciation Rate) was calculated by dividing the number of sounds in the measures by the duration (when the smallest note was 1). The group with the largest value of these weighted sum was chosen to be the right-hand part.

Finally, we selected the group corresponding to the left-hand part. The left hand often plays a chord; therefore, the group with the highest proportion of chords was selected as this part. The chord rate was calculated by dividing the number of chord sounds by the total number of sounds.

C. Correction based on restriction of electronic organ

The pitches of notes in each group correspond to particular musical instruments. In some cases, these notes cannot be played on the keyboard. Therefore, they must be corrected for compatibility with the electronic organ. To rectify the musical score, we simultaneously examined the right-hand, left-hand, and foot parts in each bar.

We initially rectified the foot-part melody. The foot part cannot play two sounds at the same time; therefore, the appearance of two sounds in the foot part must be corrected. For this purpose, we combined two sounds having the same note name into one sound. Sound combinations are based on the number of pronunciations of the musical instruments. When two notes had different sound names, we retained the sound corresponding to the root sound of the chord and erased the other sounds. When there was no sound corresponding to the root of the chord, we left a melody with a large number of pronunciation instruments. In addition, we modified the pitch range to that of the foot keyboard (i.e., approximately 1.5 octaves from C2 to G3).

Further, we rectified the right-hand and left-hand scores. The lowest and highest notes played at the same time by the right and left hands differed by more than one octave. Similar to the foot part, we rectified the sounds with the same note name if the sounds were observed to correspond to a large number of musical instruments. When two notes possessed different sound names, they were consolidated such that they never differed more than one octave. The intersection of the left and right hands indicated an infeasible playing situation. In such cases, the sound was consolidated to range from C3 to C4.

IV. EXPERIMENT

A. Experimental method

We evaluated the atmosphere of the performances before and after arranging the music and determined whether the music score was performed by considering the characteristics of the electronic organ. The existing electronic organ score maintained the musical atmosphere and was considered to be easy to play. Therefore, this score was evaluated to be the correct arrangement. On comparing the musical score of the arranged music with that of the correct answer, we were able to evaluate the similarity. In this way, we were able to judge whether the arrangement system was good or bad. Here, the evaluation value was calculated for the right hand, left hand, and foot parts.

We estimated the evaluated value using Equation (4). We subtracted the deletions (i.e., the number of missing notes), insertions (i.e., the number of lacking notes), and substitutions (i.e., the number of mistaken pitches or note names) from the total number of notes in the correct musical score. We divided this value by the total number of notes.

$$\text{The evaluation value} = \frac{\text{The total notes} - (\text{deletions} + \text{insertions} + \text{substitutions})}{\text{The total notes}} \quad (4)$$

The full score that was used for the arrangement is different from the number of parts and measures. We used the following four songs for the evaluation of our model: 1. Ave verum corpus (K.618 v, motet in D major) 2. Pictures at an Exhibition (a suite of 10 pieces having a varied promenade) 3. Salut d'Amour, and 4. Symphony No.36 "Linz" K.425 3rd movement Menuett.

B. Results

Ave verum corpus shows the result of the reduction. Figure 8 depicts the score for four measures in the song. The Figure 8(a) is the existing electronic organ score, whereas Figure 8(b) is the arranged score. The missing notes are marked in blue, and the incorrectly inserted notes are marked in red. Table I depicts the evaluation values of each song. The evaluation values of this song are 0.78, 0.73, and 0.72 for the right-hand, left-hand, and foot parts. While listening to the performance of the arranged score, the evaluation values from 0.7 to 1.0 had a good score and were arranged with the image of the original song. When the evaluation values ranged from 0.5 to 0.7, it was difficult to play using an arranged score. However, the score was arranged to maintain the image of the song. When the score was less than 0.5, it was difficult to play using the arranged score.

The evaluation values of the right-hand, left-hand, and foot parts of Ave verum corpus were more than 0.7; therefore, this is an example of how to arrange accurately. From the Table I, we can observe that the average of the evaluation values of the right-hand, left-hand, and foot part are 0.80, 0.71, and 0.77, respectively. The evaluation of each part has a value of 0.7 or more; therefore, this can be considered to be a good system. Also, this result showed that the performance was improved compared to when do not use instrumental timbre feature amount was added.

(a)

(b)

Figure 8. Arrangement results of the (a) existing electronic organ score and (b) proposed system

TABLE I. EVALUATION RESULTS OF EACH SONG

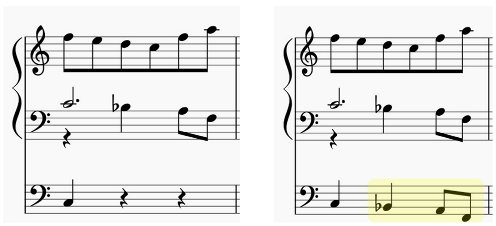
Songs name	Parts	Measures	Right hand	Left hand	Foot
Ave verum corpus	13	46	0.78	0.73	0.72
Pictures at an Exhibition	20	24	0.75	0.66	0.83
Salut d'Amour	6	99	0.81	0.67	0.85
No.36 "Linz" K.425	14	57	0.89	0.70	0.69
Averages	—	—	**0.80**	**0.71**	**0.77**

V. DISCUSSION

A. Different arrangement from the existing electronic organ score

Several measures were arranged unnaturally. If the number of musical instruments of the original score or the number of parts being played is small, only two groups may remain after clustering. If we choose a group with the foot, right-hand, and left-hand parts in that order, there will be a melody for the right hand and foot, and a musical score with no melody will be produced for the left hand. In the electronic organ, it is difficult to play using the foot part; therefore, ideally we do not play using the foot part when we compliment the melody with the right- and left-hand parts. The melody was cut off and connected unnaturally from the right to the left hand. Therefore, it was necessary to reexamine the selection method of the melody group. In the current method, we only look at different measures. We considered it better to select a melody group by considering the connection afresh using the before and after measures of the melody. In the previous method, N-gram [13] of the language model and Hidden Markov Model (HMM) [14] of the probabilistic model were used to consider connecting the melody. Using these methods, we can select an appropriate melody. We can also expect that only the left hand may have rests suddenly and that the melody will be unnaturally connected from the right to the left hand.

Further, the system performance was lowered because of the process of correcting the musical score. Additionally, a music score that was difficult to play, such as a sound that overlapped three notes, suddenly appeared in the melody of a single tone. As depicted in the yellow portion of Figure 9(b), extra sounds were inserted, and it was difficult to play using a musical score. This was because we had not considered the ease of playing instruments. It is possible to improve by considering the relation of the note numbers and the movement of fingers. In the case of Figure 9, the extra inserted sound in (b) was emitted in another part (in this case, the left-hand part) as compared with the existing electronic organ score in (a). In this way, when a mistakenly inserted sound is emitted by another part, it is better to perform processing, such as sound elimination, because the sound is supplemented by other parts.



(*a*) (*b*)

Figure 9. Example of incorrect arrangement of (a) existing electronic organ score and (b) arrangement result of the proposed system

B. Weights of feature quantities used for clustering

For each song used in the experiment, the weight of the feature quantity used for clustering is illustrated in Table II. From Table II, we can observe that the feature quantities RA of the pronunciation time, and MA of the pitch change works well. The electronic organ of the keyboard instrument is the target of this study; therefore, these features are effective. In keyboard instruments, while playing multiple melodies alone, it is difficult to play unless it is the same movement or a melody that can be pronounced simultaneously. Therefore, this system should be capable of reflecting the characteristics of the keyboard.

In future, we propose to arrange other instruments also. The results of this study depicted that we can deal with other instruments by assigning weights to the feature quantities according to the characteristics of each instrument. In case of arranging for a guitar, it would be better to increase the weight of the harmony feature quantities CA because mainly the chords are played on the guitar. Therefore, it is possible to create a guitar melody because it is possible to consider into account the clusters of chords. Furthermore, during the time of cluster selection, the musical score of the guitar can be created by choosing the melody including the chord and melody of the bass part.

TABLE II. WEIGHT USED FOR CLUSTERING

Songs name	RA	MA	CA	SR	IT
Ave verum corpus	0.3	0.4	0.1	0.1	0.1
Pictures at an Exhibition	0.3	0.3	0.2	0.1	0.1
Salut d'Amour	0.3	0.4	0.1	0.1	0.1
No.36 "Linz" K.425	0.4	0.2	0.2	0.1	0.1

VI. CONCLUSION

In this study, we proposed an arrangement system for reducing the full score of the electronic organ. First, melodies were clustered using the pronunciation time, pitch change, harmony, persistent pronunciation pattern, and timbre of instruments to summarize the melodies having similar characteristics from the full score. Next, we selected the melodies corresponding to the right-hand, left-hand, and foot parts. Finally, we modified the melodies so that they could be played using an electronic organ. In clustering, it became easier to organize the melody of musical instruments with similar timbres by considering the timbre features afresh; it also became possible to arrange according to the musical score of the electronic organ. Many errors occurred while selecting the clusters because the melody was selected only within the measures where the notes were played at the same time. In addition, the melodies were chosen without considering the connection between the before-and after-melodies. Using the N-gram or HMM, we intend to select appropriate clusters by considering the ease of connecting melodies.

In future, we plan to arrange music for other instruments also. In this preliminary step, we made an ensemble score covering musical instruments, such as the violin, saxophone, guitar, and bass. Keyboard instruments and saxophones were observed to score well. However, stringed instruments were difficult to play using the arranged score. Therefore, in our future studies, we aim to investigate the musical instrument characteristics and performance methods for other instruments and further arrange them.

REFERENCES

[1] K. Fujita, H. Oono, and H. Inazumi, "A proposal for piano score generation that considers proficiency from multiple part", IPSJ Special Interest Group on Music and Computer, 2008, pp. 47-52.

[2] S. Ito, S. Sakou, and T. Kitamura, "Automatic Arrangement of Ensemble Score by Contraction of Orchestra Score Considering Importance of Part", Trans.IPS.Japan, 2013(1), pp. 291-292.

[3] M. Matsubara et al., "Scoreilluminator: Automatic illumination of orchestra scores for readability improvement", Proceedings of the 2009 International Computer Music Conference, ICMC 2009, pp. 113-116.

[4] C. Roads, Computer music—History・Technology・Art, Tokyo Denki University Press, 2001.

[5] J. Makhoul, "Linear prediction: A tutorial review", Proceedings of the IEEE, Volume:63, Issue:4, April 1975, pp. 561-580.

[6] S. Yoshii, Digital audio processing: Tokai University Press, 1998.

[7] T. Kobayashi, "Cepstral and Mel-Cepstral Analysis of Speech", The Institute of Electronics, Information and Communication Engineers, 1998, pp. 33-40.

[8] MusicXML[Online]. http://www.musicxml.com/ [17 Mar. 2018].

[9] Musescore—Music Composition and Notation Software[Online]. http://musescore.org/ [17 Mar. 2018].

[10] Y. Dong and D. Li, Automatic speech recognition:a deep learning approach, Springer-Verlag London, 2015.

[11] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models", Acoustics, Speech and Signal Processing, 2007, pp. IV317-IV320.

[12] J. Edwards and P. Oman, Dimensional Reduction for Data Mapping-A practical guide using R, R News, Vol. 3/3, 2003, pp. 2-7.

[13] M. Tomari, M. Sato, and Y. Osana, "Automatic composition based on genetic algorithm and N-gram model", Proceedings of IEEE Conference on System, Man and Cybernetics, Singapore, 2012, pp. 202-207.

[14] T. Kathiresan, "Automatic melody generation", Master's thesis, KTH Royal Institute of Technology, 2015, pp. 25-43.