2009 11th IEEE International Symposium on Multimedia

Automatic System for the Arrangement of Piano Reductions

Shih-Chuan Chiu
National Chiao Tung University
Department of Computer Science
Hsinchu, Taiwan, ROC
scchiu@cs.nctu.edu.tw

Man-Kwan Shan
National Chengchi University
Department of Computer Science
Taipei, Taiwan, ROC
mkshan@cs.nccu.edu.tw

Jiun-Long Huang
National Chiao Tung University
Department of Computer Science
Hsinchu, Taiwan, ROC
jlhuang@cs.nctu.edu.tw

***Abstract*—Piano reduction is a process that arranges music for the piano by reducing the original music into the most basic components. In this study we present an automatic arrangement system for piano reduction that arranges music algorithmically for the piano while considering various roles of the piano in music. We achieve this by first analyzing the original music in order to determine the type of arrangement element performed by an instrument. Then each phrase is identified and is associated with a weighted importance value. At last, a phrase selection algorithm is proposed to select phrases with maximum importance to arrangement under the constraint of piano playability. Our experiments demonstrate that the proposed system has the ability to create piano arrangement.**

Keywords-piano reduction; music arrangement; algorithmic composition; phrase selection;

I. INTRODUCTION

A piano reduction is a sheet music in which an original multipart piece of music is reduced to its basic components within a two-line staff for the piano. Many famous masterpieces have a piano reduction version. The approach on how to arrange a piano reduction concentrates on how one can retain as much of what the composer has provided as possible.

In this study, we present an automatic music arrangement system for accomplishing piano reduction. In other words, the arranger does not create new counterpoints, harmonies, bass lines, voices, etc., but instead only focuses on elimination of the less important parts of the original music for the piano and keeps the new arrangement sounding like the original one. When arranging a piece of music for the piano, it is necessary to take the characteristics and the inherent limitations of the piano into account, such as pitch range and polyphonic limitation. Hence the objective is to retain as much of the original music provided as possible under the playability constraint of the piano such that the arranged piece is similar to the original.

In addition, the piano may play different roles when in an ensemble. For example, in a big band the roles played by the piano may be either solo or an accompaniment. For a solo performance, both the accompaniment and the melody should be simultaneously considered. In our proposed system, different arrangements of differing roles are also taken into consideration. The proposed approach allows users to specify the target role of the piano arrangement. To arrange different roles of a piano, we employ the concept of *arrangement elements* as proposed by Owsinski, by determining the function of a piece of music when performed by an instrument in an ensemble [7]. Good arrangement takes relationships between musical elements into consideration. Most good arrangements are limited in the number of elements that occur at the same time. There are five types of arrangement elements: *lead*, *foundation*, *rhythm*, *pad* and *fill*. The role of an instrument in an ensemble can be referred to by one of these five types of arrangement elements. In our paper, the two phrases "*role*" and "*type of arrangement element*" will be used interchangeably.

Our proposed system is designed as follows. For the purpose of role arrangement, each piece of music performed by an instrument is analyzed to obtain its type of arrangement element (role). Then each phrase in the original music is identified, and its utility is assigned according to its type of arrangement element. Finally, we select as many phrases as possible; this choice is made according to their utility and the playabilities of the piano. The new arranged music is formed by these selected phrases.

The remainder of this paper is organized as follows: In Section 2, previous studies related to music arrangement are discussed. In Section 3, the system overview is introduced first, followed by each component of the proposed system: track segmentation (3.1), arrangement element determination (3.2), phrase identification and utility assignment (3.3), playability verification (3.4) and phrase selection (3.5). Finally, the results are demonstrated in Section 4, followed by the conclusions.

II. RELATED WORK

Many works related to music arrangement focus on how to transform original music by a change in its meta-information (tempo, timbre and etc.) or content (insert note, change pitch, or re-assemble music segments and etc.) [2][6].

In the reduction technique of arrangement for an instrument, piano reduction is a two-line staff of piano music that is reduced from multipart music. The commercial software for music notation, Finale, provides a plug-in tool to combine a previous-prepared score into a two-line staff separated by a user-defined pitch value. In other words, Finale provides a tool for arranging the performance of a piano reduction rather than an automatic piano reduction system. Potter et al. presented an approach for guitar arranging [3][10]. The basic idea is to choose a set of

978-0-7695-3890-7/09 $26.00 © 2009 IEEE
DOI 10.1109/ISM.2009.105

IEEE computer society

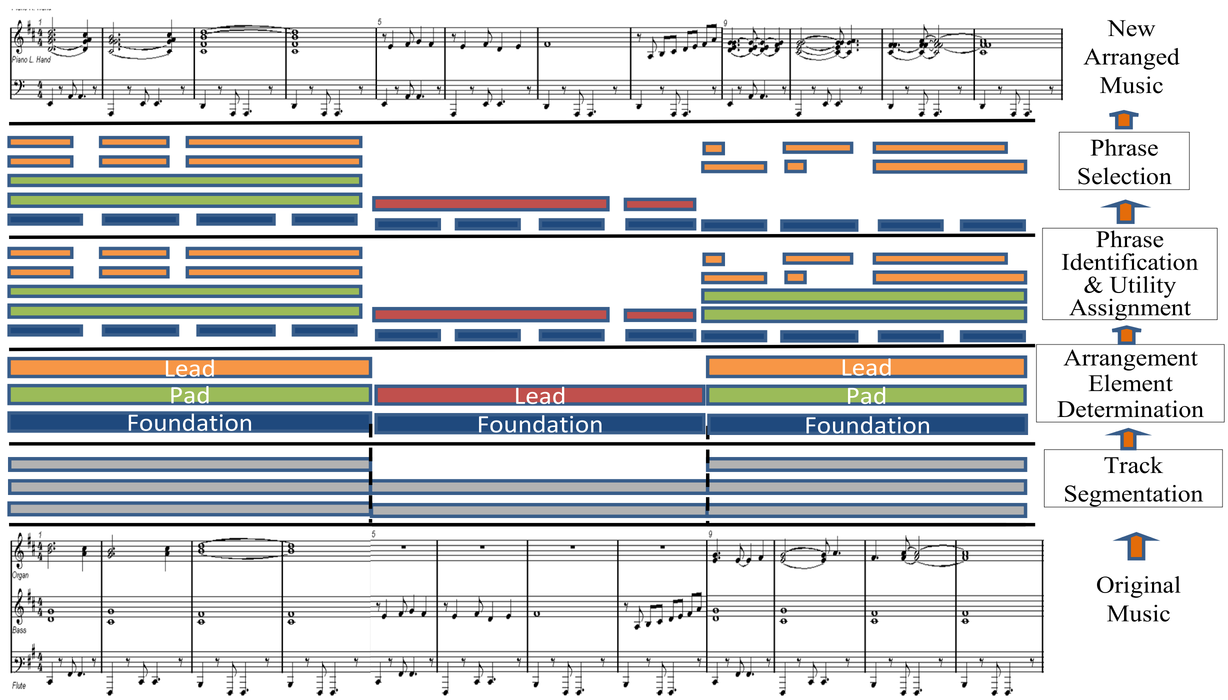

Figure 1. Illustration of proposed approach.

important notes with a constraint on the playability of the guitar by some search algorithm. However, this approach is only applicable for the solo guitar. In addition, the reduction is performed in the unit of notes rather than the phrase. If the selected notes came from different instruments, then the music meaning, such as the completeness of a piece of melody, may be lost.

III. The Proposed Approach

Given a piece of original music (multipart) and the role of the piano (proportion of five types of the arrangement element), our proposed framework outputs a playable arrangement for the piano according to the given role. Our arrangement framework consists of four main steps; these are illustrated in Fig. 1. First, the original music object is partitioned into *segmented tracks*. Next, the type of arrangement element of each segmented track is determined by a classifier. The classifier is constructed by machine learning from expert-annotated segmented tracks. In the third step, for each segmented track, the phrases are identified by *phrase identification*. A utility value is assigned to each phrase according to the corresponding type of arrangement element. Finally, parts of phrases are selected by a *phrase selection* algorithm based on the associated utility and playability of the piano. The new arranged music is formed by these selected phrases.

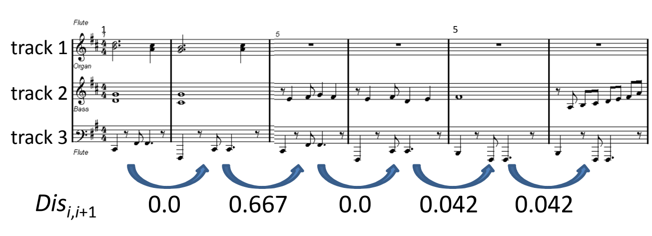

Figure 2. An example of section segmentation.

A. Track Segmentation

A track performed by an instrument may play different roles of the arrangement elements. For example, a violin, demonstrating the pad element may change to lead. Therefore, in our approach a track is segmented into *segmented tracks* first. A segment track is a period in which all the instruments remain consistent in the roles of arrangement elements. For example, in Fig. 1, the original music that consists of three tracks is segmented into three periods. It is observed that a time point, where many instruments stop and others start, is very likely to be a cut point. According to this observation, we define the dissimilarity function between consecutive measures as follows.

$$Dis_{i,i+1} = \frac{\sum_{t} \left| NSBM_{i,t} - NSBM_{i+1,t} \right|}{TotalNumOfTrack \times BeatPerMeasure} \quad (1)$$

where $NSBM_{i,t}$ is the number of non-silent beats in the i-th measure of track t; *TotalNumOfTrack* is the total number of tracks; and *BeatPerMeasure* is the number of beats per measure. We define a threshold value τ to decide cut points. If the dissimilarity between measure i and i+1 is larger than τ, then there is a cut point between them. The threshold τ is heuristically designed to be 0.5. If $Dis_{i,i+1}$ is larger than τ, it means more than half of the instrument roles change between measure i and $i+1$.

Fig. 2 gives an example that displays the dissimilarity for each successive measure in this music. For the dissimilarity between the 2-th and 3-th measures, $Dis_{2,3}$ is computed as follows. In the music example, the time signature is 4/4 with 3 tracks; that is, each measure has 4 beats (*BeatPerMeasure*=4, *NumTrack*=3). In the first track, there are four beats in measure 2 ($NSBM_{2,1}$=4), and there is no sound in measure 3 ($NSBM_{3,1}$=0). The other measures ($NSBM_{2,2}$, $NSBM_{3,2}$, $NSBM_{2,3}$, $NSBM_{3,3}$) can be calculated in the same way. Thus, $Dis_{2,3}$=(|4-0|+|0-3.5|+|3-3|)/(4×3) = 0.667. Since 0.667 is larger than 0.5, there is a cut point between measure 2 and 3.

B. Arrangement Element Determination

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

According to [7], there are five types of arrangement elements: *lead*, *foundation*, *rhythm*, *pad* and *fill*.

Lead - a melody part and its counterpoint. It is usually a lead vocalist or solo instrument.

Foundation - the main rhythm in a musical piece. The foundation is always a regular pattern played by a drum (especially bass drum or snare) or bass instrument.

Rhythm – the broken beat is counted to the foundation element played by any instrument. It is more complicated in beat and used to increase music fluency.

Pad - consists of a long, sustained note or chord. It is usually played by a string, organ or synthesizer. Generally, the pad can also denote those sounds which create ambiance.

Fill - usually appears in the spaces between the lead lines. It is similar to a conversation. While the lead acts as a caller, the fill acts as a respondent.

To determine the type of arrangement element for a given segmented track, we utilize a five-way classification technique to classify each segmented track into one of the five classes: foundation, rhythm, pad, lead and fill.

One of the important steps of the classification technique used here is to decide which features are to be extracted and, more importantly, which features accurately represent the segmented track. The properties of the piano are an important factor in influencing the assignment of its arrangement element. Moreover, the role of a segmented track highly depends on those of the other segmented tracks in the music, especially parallel tracks. Thus, the features we consider are not only global features (common features) but are also local features (related to the other segmented tracks). The extracted features and corresponding descriptions are listed in Table 1.

The classification algorithm we adopt is a support vector machine (SVM). SVMs construct a separating hyper-plane using *support vectors* (a subset of training data) that maximizes the margin between two data sets. A "good" separation is achieved when the hyper-plane has the largest distance to the neighboring data points of both classes. After the hyper-plane is determined (training phrase), the SVM model can answer or predict the class of a new given example. We use SVM to construct the classifier learned from expert-annotated tracks. Given a segmented track of unknown type, the constructed classifier will output the probability distribution over five types of arrangement elements. Because SVM is a binary classifier, $C_2^5 = 10$ SVMs are employed for five-class classification. To obtain class probabilities, logistic regression models are then used to fit to the outputs of the SVM. In the multi-class case, Hastie and Tibshirani's pairwise coupling method is employed to estimate the probabilities.

C. Phrase Identification and Utility Assignment

After track segmentation and determination of arrangement element types, the next step is to identify the phrases from each segmented track and to assign a utility to each phrase. The musical phrase is the basic unit for an arrangement in our proposed approach.

TABLE I. FEATURES FOR THE CLASSIFIER.

Parameter	Type	Description
AvgPitch	G	Average pitch in the segmented track
AvgDuration	G	Average duration in the segmented track
DevPitch	G	Pitch deviation in the segmented track
IsPercussionChannel	G	Is Percussion Channel (usually channel 10)
PolyphonicRate	G	Proportion of note occurring in the same time
AvgPitchRank	L	Rank of average pitch in parallel segmented track
AvgDurationRank	L	Rank of average duration in parallel segmented track
IsHighestPitchPart	L	Is the segmented track with the highest average pitch in parallel segmented tracks
IsLowestPitchPart	L	Is the segmented track with the lowest average pitch in parallel segmented tracks

G: global feature, L: local feature

First, the monophonic lines are extracted using the approach proposed in [4] to preserve the best lead voice. In [4], the concepts of top notes and base notes were proposed as a method in identifying the base lines. An example of a top note is the highest pitch in a chord while that of a base note is its lowest pitch. The baseline is therefore formed by the base notes. For each identified monophonic baseline, we employ the local boundary detection model (LBDM) [1] to segment the baseline into phrases. LBDM extracts the pitch interval sequence, the inter-onset interval sequence and the reset sequence from the main melody. Then these three sequences are integrated into the sequence of boundary strength values measured by the change rule and the proximity rule. The resulting peaks of the boundary strength

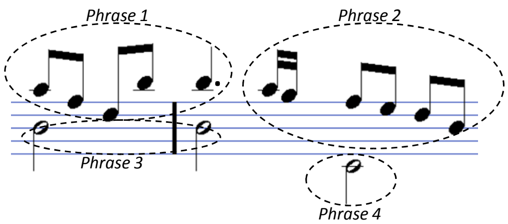

Figure 3. An example of phrase identification.

value sequence are regarded as the phrase boundaries. Fig. 3 shows an example of phrase identification.

While the objective of piano reduction is to keep as much of the original music as possible so that the resulting arrangement is similar to the original, each phrase is associated with a so-called *utility*, which is used to indicate its specific importance within the arrangement. The utility $U(p)$ of a phrase p is defined as follows:

$$U(p) = \mathrm{Pr_r}(p) + \mathrm{E}(p) \tag{2}$$

where $\mathrm{Pr}_r(p)$ is the probability that phrase p belongs to the user-specified target role r, and $\mathrm{E}(p)$ is the entropy of p. The first factor of utility concerns only the role of phrase p, whereas the second factor specifically concerns the richness of phrase p. As mentioned before, the probability distribution of each segmented track over five types of arrangement elements is obtained by the step of the arrangement element. The probability distribution of a phrase is inherited from the segmented track in which it is located. The entropy of phrase p measures the randomness of distinct pitch values in p.

D. Playability Verification

Several studies have been devoted to the research of automatic piano fingering [5][13]. However, these works cannot be used to determine whether a piece of music can be played using the piano.

To verify the playability of a piece of music for the piano, the following rules are considered.

Rule 0: At a time, at most five phrases can be played simultaneously. This is because the hand of a piano player has only five fingers .

Rule 1: In the piece of music, each pitch is within the pitch range of the piano, which ranges from the A three octaves below middle C to the C four octaves above middle C.

Rule 2: In each set of notes played simultaneously, the distance between the highest pitch and the lowest pitch cannot exceed the reach of the fingers of each hand, such as 14 semitones.

E. Phrase Selection

The last step of the proposed piano reduction is to select as many phrases as possible according to their utility and the playabilities of the piano. Given a set of n phrases, $P = \{p_0, p_2, \ldots, p_{n-1}\}$ with start times, end times and utilities $\{U(p_0), U(p_2), \ldots, U(p_{n-1})\}$, the objective of phrase selection is to find a *feasible* subset, S of P, such that the sum of the utilities of

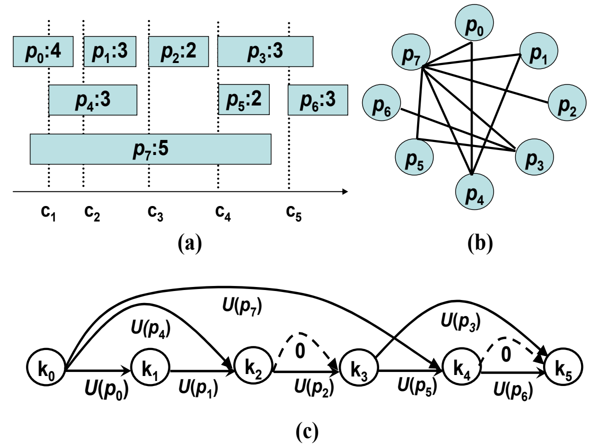

Figure 4. An illustration of phrase selection. (a) the set of phrases (b) interval graph (c) selection graph.

the selected phrases in S is maximized. A set of phrases is feasible if it agrees with the playability rules.

To solve this optimization problem, a straightforward approach would be to enumerate and verify all possible combinations of selected sets, i.e., the power set. In other words, each phrase is either selected or ignored. Therefore, the time complexity is up to $O(2^n)$.

In this paper, as space is limited, we only exhibit the algorithm to find the optimal solution that agrees with the zeroth rule of playability. Given a set of n phrases, this algorithm first constructs an *interval graph* in which there is a vertex for each phrase and an edge between two vertices if the corresponding phrases overlap. Two phrases are overlapped if the intersection of their interval is nonempty.

Then, all the maximal cliques of the constructed interval graph are identified and are ordered linearly according to the start time of the occurrence. For example, Fig. 4 illustrates an example of phrase selection. In this example, there are seven phrases associated with utility values. The interval graph constructed from the seven phrases is shown on the right. Five maximal cliques, c_1, c_2, c_3, c_4, and c_5, are identified from this interval graph.

Having identified the maximal cliques, a *selection graph* is constructed. The selection graph is a directed graph in which there is a vertex k_s for each maximal clique c_s. Moreover, a dummy vertex k_0 is created as the source vertex. For each phrase p_i in cliques $c_s, \ldots, c_{s+t}$, an edge with weight $U(p_i)$, the utility of , is created from the vertex k_{s-1} to k_{s+t}. If the size of a clique c_s is not the maximum, a dummy edge is added from vertex k_{s-1} to k_s with weight zero. The zero-weighted edge is there to ensure that there exists a connected path from the source to the destination.

At last, the algorithm makes its phrase selections by repeating the following steps five times:

Step 1: Find the longest path from the source to the vertex corresponding to the last clique. Every phrase corresponding to the edges along this longest path is chosen.

Step 2: Modify the selection graph by removing the edges in the longest path in the first step, with the exception of dummy edges.

For example, in Fig. 4, if the number of fingers is limited to two, the algorithm will find the longest path corresponding to the set of phrases, $\{p_0, p_1, p_2, p_5, p_6\}$ first. The next iteration will find the path with respect to the set of phrases, $\{p_3, p_4\}$.

IV. EXPERIMENTS

Our music arrangement system was implemented in Java along with two open source software packages, jMusic [9] and Weka [8]. jMusic is an environment for manipulating MIDI data, and Weka is a provider of machine learning tools. We chose MIDI-format music as a source of symbolic data.

A. Accuracy of Arrangement Element Determination

We collected the segmented tracks by performing track segmentation for each music object in our music database first. Then the music-trained experts were asked to annotate the type of arrangement element for some of the segmented track. To annotate the type of arrangement element, the musically trained experts were asked to read the book on music arranging, [2], to learn about the arrangement elements. There are 240 segmented tracks annotated in total. Among them, 78, 56, 15, 67, and 24 belong to foundation rhythm, pad, lead and fill, respectively. The confusion matrix of arrangement element determination is shown in Table 2. The f-measures are 0.907, 0.826, 0.72, 0.813 and 0.4, respectively. The results are acceptable except for the fill. One explanation for this is that fill, which usually appears between lead lines, is very similar in perception to lead.

TABLE II. CONFUSION MATRIX FOR ACCURACY OF ARRANGEMENT ELEMENT DETERMINATION.

	Foundation	Rhythm	Pad	Lead	Fill
Foundation	73	3	0	2	0
Rhythm	7	45	0	3	1
Pad	0	3	9	2	1
Lead	3	1	0	61	2
Fill	0	1	1	15	7

B. Effectiveness of Piano Reduction

It is difficult to evaluate the effectiveness of our system because the evaluation of effectiveness in works of art often comes down to subjective opinion. In 2001, M. Pearce proposed a method to evaluate the computer music composition system [8]. We adopted this method to design experiments.

The proposed system can be considered successful if the subjects cannot distinguish the system-arranged music from the human-arranged music. Twenty-two graduate and undergraduate students including four well-trained music experts were invited to act as subjects. The prepared dataset consisted of 8 human-arranged and 8 system-arranged music objects. These music objects were sorted randomly and displayed to the subjects. The subjects were asked to listen to each music object, and to determine whether it is system-arranged or human-arranged music. The proportions of correctly discriminated music were calculated from obtained result (Mean is the average of the accuracy). The significant test was performed with the one-sample t-test against hypothesized value 0.5 (the expected value if subjects discriminated randomly).

The result is shown in Table 3. We accept the hypothesized value of 0.5 using the 0.05 level of significance for a student's t-test. This implies that it is difficult to discriminate between system-arranged music and human-arranged music. However, the p-value is not statistically significant due to the small number of subjects. The discrimination rate from all the subjects except for the experts is much closer to 0.5 than the discrimination rate from all the subjects. As we expected, the experts can discriminate much more precisely.

TABLE III. THE RESULTS OF THE DISCRIMINATION TEST.

	Mean	SD	DF	t	P-value
All subjects	0.443	0.138	21	-1.94	0.066
All subjects except experts	0.444	0.15	17	-1.61	0.1258

SD: the standard deviation; DF: the degree of freedom; t: t statistic.

C. Case Study

We demonstrate a system-arranged piano reduction. Fig. 5a and Fig. 5b show an excerpt from measure 1 to 8 of the original song and the new arrangement, respectively. The original arrangement contained five instruments: three acoustic guitars (one for melody and two for chord), one bass and one violin. We set the parameters [Foundation, Rhythm, Pad, Lead, Fill]=[0,0,0,1,1] for right hand and [1,1,1,0,0] for left hand. We can see that the system has an ability to select the correct melody for right hand. This is because the arrangement element determination contributed to segmented tracks that assigned proper utility to those phrases. In phrase selection, the system maximized the total utility consideration and allowed five phrases that overlap, such that not only the phrase of the main melody, but also the other phrases were selected. For our left-hand part, the phrases in the bass and harmonic voice were also maximized for the left hand.

Also note that the type of arrangement element of the bass line belongs to the foundation. In the arranged results, the bass note is absent for the following two reasons. First, the bass note is a one-note phrase and far away from the other notes. If it is selected, no other phrase can be selected according to the playability verification. Second, the utility of this bass-note phrase is not high enough to discard the other phrases. We think this song was arranged successfully for a piano reduction. It was also playable. In Fig. 5c, a piano reduction for accompaniment was generated from the same song. Compared to the parameter setting for solo, we set [1,1,1,0,1] and [1,0,0,0,0] for right and left hand, respectively. The phrases for accompaniment could be selected due to the correctness of our arrangement element determination method.

V. Conclusions

We presented here a system that arranges multipart scores for piano with consideration of the role played by instrument. While our system is able to produce viable and adaptable arrangements for the piano, it could be applied to many other instruments, with modification of the playability function used here. The different roles of arrangement in an ensemble can also be successfully adapted to our system. In future work we will extend our system to arrange for different instruments or an entire ensemble.

References

[1] E. Cambouropoulos: "The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing," *Proceedings of the International Computer Music Conference*, 2004.

[2] J. W. Chung: "The Affective Remixer: Personalized Music Arranging," *Proceedings of ACM International Conference on Computer-Human Interaction,* 2006.

[3] R. Daniel and W. D. Potter: "GA-based Music Arranging for Guitar," *Proceedings of International Congress on Evolutionary Computation,* 2006.

[4] S. Lui, A. Horner, and L. Ayers: "'MIDI to SP-MIDI Transcoding Using Phrase Stealing" *IEEE Multimedia*, Vol. 13, No. 2, 2006.

[5] A. A. Kasimi, E. Nechols and C. Raphael: "A Simple Algorithm for Automatic Generation of Polyphonic Piano Fingerings," *Proceedings of International Conference on Music Information Retrieval,* 2007.

[6] T. Nagashima and J. Kawashima: "Experimental Study on Arranging Music by Chaotic Neural Network," *International Journal of Intelligent Systems,* Vol. 12, No. 4*,* 1997.

[7] B. Owsinski: *The Mixing Engineer's Handbook*, Thomson Course Technology, 1999.

[8] M. Pearce and G. Wiggins: "Towards A Framework for the Evaluation of Machine Compositions," *Proceedings of Symposium on Artificial Intelligence and Creativity in the Arts and Sciences,* 2001.

[9] A. Sorensen and A. R. Brown: "Introducing jMusic," *Proceedings of the Australasian Computer Music Conference*, 2000.

[10] D. R. Tuohy and W. D. Potter: "'An Evolved Neural Network/HC Hybrid for Tablature Creation in GA-based Guitar Arranging," *Proceedings of the International Computer Music Conference*, 2006.

[11] G. White: *Instrumental Arranging*, McGraw-Hill, January 1992.

[12] I. H. Witten and E. Frank, Data Mining: *Practical Machine Learning Tools and Techniques*, CA: Morgan Kaufmann, 2005.

[13] Y. Yonebayashi, H. Kameoka and S. Sagayama: "Automatic Decision of Piano Fingering Based on Hidden Markov Models," *Proceedings of International Joint Conference on Artificial Intelligence,* 2007.

Figure 3. (a) Original music: an excerpt from an Irish folk song "Green Grow the Lilacs"
(b) Piano reduction for solo (c) Piano reduction for accompaniment.