A COMPARISON OF STATISTICAL AND RULE-BASED MODELS FOR STYLE-SPECIFIC HARMONIZATION

Ching-Hua Chuan
University of North Florida
School of Computing
Jacksonville, FL
c.chuan@unf.edu

ABSTRACT

The process of generating chords for harmonizing a melody with the goal of mimicking an artist's style is investigated in this paper. We compared and tested three different approaches, including a rule-based model, a statistical model, and a hybrid system of the two, for such tasks. Experiments were conducted using songs from seven stylistically identifiable pop/rock bands, and the chords generated by the systems were compared to the ones in the artists' original work. Evaluations were performed on multiple aspects, including calculating the average percentage of chords that were the same and those that were related, studying the manner in which the size of the training set affects the output harmonization, and examining a system's behaviors in terms of the ability of generating unseen chords and the number of unique chords produced per song. We observed that the rule-based system performs comparably well while the result of the system with learning capability varies as the training set grows.

1. INTRODUCTION

Automatic generation of harmony is a natural extension and application of harmonic analysis, an essential component in music information retrieval. Previous research in automatic harmonization focuses on Western classical music, applying various techniques ranging from rule-based models [4] to genetic algorithms [10] in order to automate the process of harmonization in styles. An example would be the four-part harmonization in the Baroque period. Recently, systems have been developed for automatic harmonization in popular music [3, 7, 9], i.e., creating a sequence of chords for a given melody representing the vocal part in a song. However, the concept of style is loosely defined or even missing in most of these systems. As the Beatles represents a firmly defining role in pop/rock music, the style of the individual artist must be considered.

In this paper we compare three different approaches for style-specific harmonization in popular music. The three approaches demonstrate a wide spectrum of techniques: a knowledge-driven model, a data-driven model, and a hybrid system combining the two. We conducted experiments by taking the melody of songs from seven identifiable pop/rock bands as the input for the three systems, and compared the system-generated chords with the ones in the original artists' work. For systems with learning capabilities, we analyzed the relationship between the size of the training set and the quality of the output harmonization. We also examined the characteristics of each system in terms of the number of unique chords it generates for each song, and its ability to produce chords that are not included in training sets.

2. PROBLEM DEFINITION

Suppose a melody consists of m monophonic notes, $\{a_1, \ldots, a_m\}$, harmonized by a sequence of n chords $\{C_1, \ldots, C_n\}$, $1 \leq n \leq m$. The melody can also be represented as a set of n melody segments, $\{M_1, \ldots, M_n\}$, and each of the segments contains notes harmonized by a particular chord. For example, the melody segment M_i, harmonized by the chord C_i, can be represented as:

$$M_i = \{ a_{(\sum_{j=1}^{i-1}|Mj|)+1}, \ldots, a_{(\sum_{j=1}^{i-1}|Mj|)+|Mi|} \}, \tag{1}$$

where $|M_j|$ is the number of notes in the melody segment M_j. The location of a chord often aligns with the bar line between two measures, but not necessarily, as more than one chord may appear in a bar. Chords for two adjacent melody segments may be identical or different.

In order to generate chords for a given melody, the harmonization task requires two steps: segmenting the melody into melody segments and selecting a chord for each melody segment. In this paper we focus on the second step, chord selection, and assume the information about segmentation is given. Each chord C_i is selected among 24 candidates, 12 major triads and 12 minor ones. The choice of the

24 triads is partially due to the fact, as indicated in [8], that 96% of the chords in the collected work by the Beatles are major and minor triads. And the choice of triads is also because of our intention to focus on the fundamental chords.

3. SYSTEMS

3.1 The Rule-Based Harmonic Analyzer

The Harmonic Analyzer [11] (HA) proposed by Temperley and Sleator applies preference rules to rhythm analysis and harmonization in the Western classical music tradition. To harmonize a melody, the system first divides it into segments, and then assigns the root of the chord for each segment, without indicating the mode (major or minor). For the purpose of this paper, we focus on the process of root finding. The system operates on the application of the four Harmonic Preference Rules (HPR):

HPR 1 (Compatibility Rule): prefer certain TPC (tonal pitch-class)-root relations over others, in the following order: $\hat{1}$, $\hat{5}$, $\hat{3}$, $b\hat{3}$, $b\hat{7}$, $b\hat{5}$, $b\hat{9}$, ornamental;

HPR 2 (Strong Beat Rule): prefer chord-spans that start on strong beats of the meter;

HPR 3 (Harmonic Variance Rule): prefer roots that are close to the roots of nearby segments on the line of fifths;

HPR 4 (Ornamental Dissonance Rule): prefer ornamental dissonances that are (a) closely followed by an event a step or half-step away in pitch height, and (b) metrically weak.

Given a melody segment, a score is calculated for each of the possible 12 roots as a weighted sum using the four preference rules. The compatibility rule (HPR 1) assigns a score to each note in the melody segment depending on the relationship of the note to the root. If the note is the tonic ($\hat{1}$) of the root, it receives the highest score. Notes that are not listed in the compatibility rule are given penalties, depending on the inter-onset interval between the note to the next note a step or half-step apart in pitch and the note's metrical strength (HPR 4). Whenever a new root is selected for a segment, i.e., a chosen root is different from the one in the previous segment, it receives a penalty based on the strength of the beat where the new root starts. If the new root starts at a strong beat, it will receive a lower penalty (HPR 2). To apply the harmonic variance rule (HPR 3), a center of gravity is calculated as the average position of roots in all previous segments on the line of fifths, weighted by the length and how recent the segments are. The current root is then assigned a penalty based on its distance to the center of gravity. The scores calculated on HPR1 and HPR 3 are further weighted by the length of the segment. Finally, a dynamic programming algorithm is applied to retrieve the path of roots that report the highest overall score.

We used the implementation of the system provided by Temperley and Sleator [12] for comparison in this paper. We converted melodies in the MIDI format to text files containing a sequence of note events with beat structures as the required input for the HA system. In order to make the output of the HA system comparable to the ones from other systems, we expanded the output root into a major or a minor triad. We interpreted the chords as being the common ones as described in the textbook for Music Theory [6]. The common chords, written in Roman Numerals, include I, ii, iii, IV, V, vi, and vii. For example, when a root G is reported by the HA system in a song in the key of C major, we assign a G major (V) instead of a G minor (v) chord. For a root not listed as either major or minor in the set of common chords, we randomly assign a mode to the root.

3.2 Hidden Markov Models

Statistical approaches, particularly Markov Models, have been commonly utilized for harmonic analysis and generation in Western classical music [1, 5]. More recently, MySong [9] uses HMMs to automatically choose chords to accompany a vocal melody. Five categories of triads are considered in the MySong system, including major, minor, augmented, diminished and suspended triads. Chords are represented as their functional roles in relation to the key, which is given along with each song. The system models two types of relations: the co-occurrence of a chord and the distribution of pitches in the melody segment, and the co-occurrence of two chords observed adjacently. Two probability matrices are constructed to record the statistical information about the two relations. The first matrix, melody observation matrix, records duration-weighted melodic pitch class histogram observed in training examples for all the chords in consideration. The second matrix, chord transition matrix, shows the logarithmic likelihood of the transition from one chord to another observed in the training examples. To generate chords for an input melody, a pitch class histogram is first produced for each melody segment as the observed state, and the likelihood of a chord chosen for that melody segment is calculated using melody observation matrix. Combining the resulting logarithmic likelihood with chord transition probabilities, the Viterbi algorithm is then applied to retrieve the most likely possible chord sequence for the entire melody.

The main design goal of MySong is different from the topic concerned in this paper. The system was trained on hundreds of songs by various artists across many genres at once, without concentrating on any particular style. The final chord sequence was controlled by users through the adjustment of two options: "happy factor" generates more major triads, while "jazz factor" assigns more weights on the melody observation matrix than on the chord transition matrix. Regardless of the different design goal, the underlying HMMs in MySong can be easily adapted to the generation of style-specific harmonization with proper modifications. Inspired by MySong, we implemented a HMM-based model for style-specific harmonization. We maintained the two matrices and the way they were calculated, and also applied the Viterbi algorithm to retrieve the final chord sequence. However, we discarded the two user options with the result that the generated chord sequence completely depends on the statistical information observed in the training examples. We also limited chord selection to among major and minor triads only, resulting in a 24-by-12 melody observation matrix and a 12-by-12 chord transition matrix. Information such as melody segment and key is given. During the process of training, only songs written by one artist or band are supplied.

It is important to discuss the differences between the rule-based HA system and the HMM approach. In addition to the basic musical terms such as pitch, pitch class, chord and key that exist in both systems, the HA system has embedded more knowledge of abstract musical structures, including scale, rhythmic hierarchy, ornamental and circle-of-fifths. The functional role of each melody note and that of each chord in relation to the hierarchical and abstract structure of the song are well defined in the HA system as preference rules. To generate harmonization for a given melody, chords are selected by a series of calculations using predefined scores and penalties. In contrast, none of these abstract structures are considered in the HMM approach. Only two relations are modeled in the HMM system: pitch class distribution in melody for each segment (the observed state) and transitions between adjacent chords (transitions between states). The preference of such relations in HMM is completely determined by the training examples without using any pre-set scores or penalties.

3.3 Automatic Style-Specific Accompaniment System

In [3], Chuan and Chew proposed an Automatic Style-Specific Accompaniment (ASSA) system that generates accompaniments in a particular style to a melody given only a few training examples. The system takes a hybrid approach, applying statistical learning on top of a music theoretic framework. In ASSA, the relation between melodic notes and chordal harmonies is modelled as a binary classification task called chord tone determination: if the note is part of the chord structure, then the note is classified as a chord tone; otherwise it is labelled a non-chord tone. Each melody note is represented using 73 attributes, including pitch, duration, metrical strength, its relation to the neighbouring tones, phrase location, etc. These attributes describe the functional role of each melody note in the various abstract musical structures of the song. However, unlike the HA system, the preference or suitability of a certain type of note or chord is not pre-programmed into the system; it is learned from the training examples. Therefore, the resulting classifier, a trained decision tree in ASSA, is completely determined by the style shared in common by the training songs.

Instead of representing chord transitions as pairs (source chord and destination chord) as in the HMM approach, the ASSA system applies neo-Riemannian transforms [2] to focus on the musical relationship between the two chords involved in the transition and the movements of pitches from one chord to another. For example, a transition from a C major triad to an E minor triad is described using the leading tone exchange (L) operation[1] because the two triads share the pitches e and g, but the pitch c in C major is replaced by the E minor's pitch b, which is the leading tone in C major. The transition from F major triad to A minor triad is also described using the same L operation, while such transition is recognized as a different chord pair (C major, E minor) in the HMM approach. Chord transitions in the ASSA system are represented in a manner that reflects their relation on the circle-of-fifths and voice leading between the chord tones. But unlike the HA system, which always prefers the movement in the shortest distance on the circle-of-fifths, the applicability of the transition type is determined by the training examples.

Another difference between ASSA and the previous two systems can be observed in the generation of the final chord sequence for harmonization. ASSA generates harmonization in a divide-and-conquer fashion. The system first divides the input melody into sub-phrases delineated by bars in which melody notes strongly imply triads; then it generates a sequence of chords for each sub-phrase independently. For each sub-phrase, a Markov model is used to calculate probabilities of all possible chord series. Given a series of n chords, $\{C_1, \ldots, C_n\}$, where each chord is indexed by its segment number, the probability that this chord series occurs can be expressed as:

$$\begin{aligned} & P(C_1,\ldots C_n \mid S_1,\ldots S_n) \\ & = P(C_1 \mid S_1)P(C_2 \mid C_1, S_1, S_2)\ldots P(C_n \mid C_{n-1}, S_{n-1}, S_n) \\ & = P(C_1 \mid S_1)P(NRO_{1,2} \mid S_1, S_2)\ldots P(NRO_{n-1,n} \mid S_{n-1}, S_n), \quad (2) \end{aligned}$$

[1] The four fundamental operations in neo-Riemannian transforms are I (Identify), L (Leading-tone exchange), P (Parallel) and R (Relative).

where $NRO_{i-1,\,i}$ is the neo-Riemannian operation between chord C_{i-1} and C_i, and S_i is the phrase position of segment *i*, which falls into one of four possible categories: start, middle, ending and final. These sub-phrases of chords are at last combined, with refinements, to produce the chord progression for the entire melody.

4. EXPERIMENTS AND RESULTS

4.1 Experiments

The objective of the paper is to examine the effectiveness of the three approaches – a rule-based system, a statistical model and a hybrid system – for the automatic generation of style-specific harmonization. We used 140 songs by seven stylistically distinct pop/rock bands, including the Beatles (B), Bon Jovi (BJ), Green Day (GD), Guns N' Roses (GR), Indigo Girls (I), Keane (K) and Radiohead (R). Songs by the same band are considered to have similar styles. We obtained information about each song such as melody, chord and key from the commercial lead sheet. Melodies were encoded in the MIDI format while chords and keys were written in text files with melody segments specified.

For systems with learning abilities, we conducted the Leave-One-Out test. We selected one song as the test song and formed a training set using the remaining songs by the same artist. We then compared the generated chords with the ones given in the commercial lead sheet (the ground truth) of the test song. To examine the manner in which the number of training examples affects the performance of the systems, we constructed training sets with various sizes by gradually adding one song into the set. Suppose we have *m* songs by an artist and *n* represents the number of songs in the training set, $1 \leq n \leq m$-1. For each test song, we can construct C_n^{m-1} different training sets. Therefore, for each *n*, we will have results from *m* x C_n^{m-1} different test instances. The number of test instances grows quickly and becomes infeasible as *m* and *n* increase. For example, if we have 20 songs by an artist and we form test sets of 10 songs, the resulting number of test instances is 20 x C_{10}^{19} = 1847560. We limited the number of training sets by randomly choosing 120 training sets for each test song if the total number of possible training sets exceeds 120. Therefore, for each *n*, the number of test instances is bounded by 120 x *m*. On the other hand, for the rule-based HA system that does not require training examples, the total number of test instances for an artist is equivalent to *m*.

4.2 Results

4.2.1 Same Chord Percentage

Figure 1 shows the average percentage of generated chords that are identical to the ones in the ground truth with 95% confidence interval. Notice that the ASSA system reports a higher same chord percentage when the number of training songs increases. But the same chord percentage of HMM decreases as the increment of training songs increases in all cases except the one shown in Figure 1 (b). In general, ASSA reports higher or at least equivalent same chord percentage as HMM. However, comparing with ASSA and HMM, it is difficult to make general comments on the result of rule-based HA (the one with zero training songs) because of its wide confidence interval.

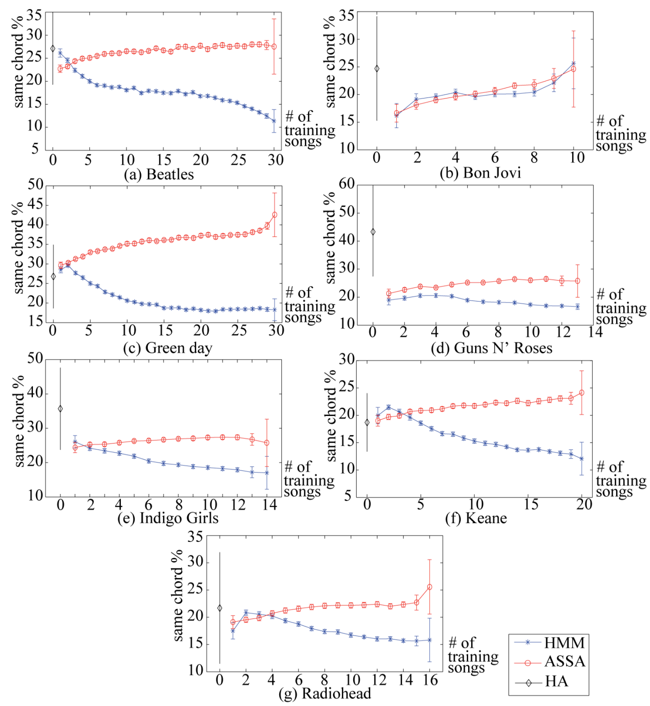

Figure 1. Same chord percentage with different sizes of training sets.

4.2.2 Related Chord Percentage

Figure 2 shows the average percentage of generated chords that are closely related to the ground truth. Two chords are considered closely related if they show one of the following relations: identical, dominant, subdominant, relative, parallel, dominant/relative, dominant/parallel, subdominant/relative and subdominant/parallel. For example, if the ground truth is C major, the closely related chords in the order are C major, G major, F major, A minor, C minor, E minor, G minor, D minor and F minor. When related chord percentage is considered, the rule-based HA performs the best in general. HMM and ASSA perform similarly, but as the number of training songs increases, the results for ASSA improves while those for HMM decline.

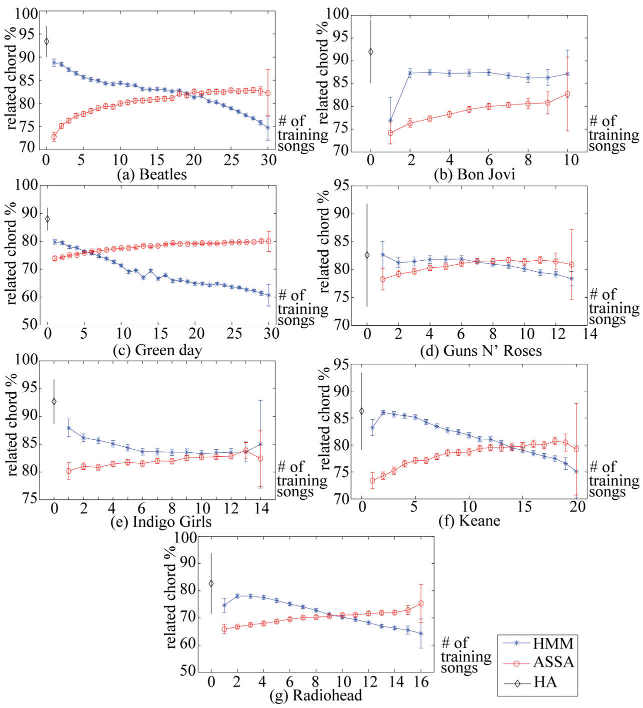

Figure 2. Related chord percentage with different sizes of training sets.

4.2.3 Average Number of Unique Chords

We also examine the number of unique chords generated for each song by the three systems, and compare that with the number of unique chords in each band's original songs. Each unique chord chosen by a composer is analogous to a color used by a painter, and the number of colors that appear in a painting is usually considered a contributing factor of a painting style. The number of unique chords equals the total number of chords in the sequence subtract the number of duplicate chords. Figure 3 shows the average number of unique chords generated by the three systems and in the original songs. Notice that the average number of unique chords generated by the rule-based HA system is the closest to but slightly lower than the ground truth (GT). The number of unique chords generated by HMM grows as the number of training examples increases, which provides more chords as cases for HMM to learn from. In contrast, the number of unique chords generated by ASSA drops and becomes closer to the ground truth when the number of training examples increases. This may result from the use of the neo-Riemannian transform, which only represents the relative relation in the transition between chords, allowing more freedom to choose chord pairs that are not included in the training set as long as they share the same transition.

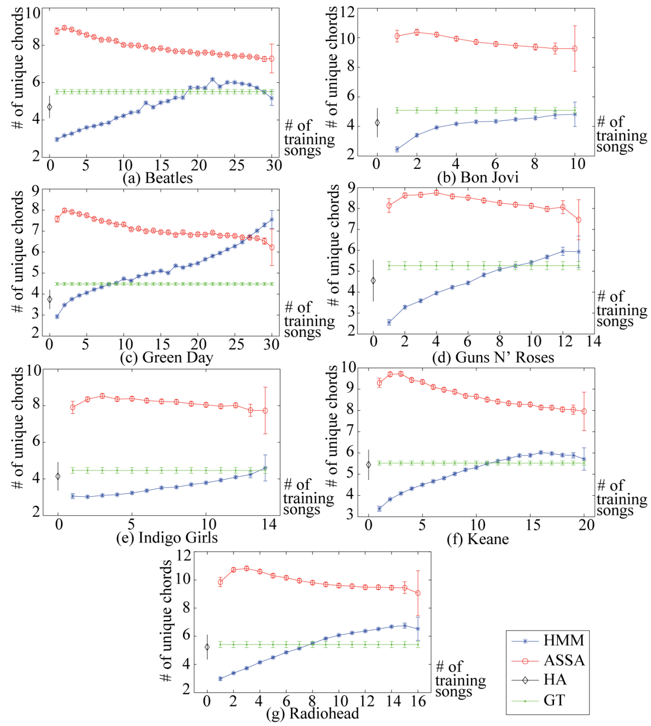

Figure 3. Average number of unique chords per song using HMM, ASSA, HA and in the ground truth (GT).

4.2.4 Average Number of New Chords

For systems that require training examples, it is important to study how these examples affect the output. Particularly, we are interested in the system's ability to generate chords that are not given in the training examples. For comparison, we also investigate the original songs to observe the number of chords in a song that do not appear in a given set of other songs by the same artist. We label these unseen chords as new chords.

Figure 4 presents the average number of new chords generated by HMM and ASSA, and in the original accompaniment, the GT. In the original accompaniments, when the training set is small, there are always one or two new chords in each song. As the training set grows, the training examples gradually cover all the chords in each song. In ASSA, because of the neo-Riemannian framework, it demonstrates the ability to create new chords but tends to generate too many when the training examples are too few. More training examples help ASSA become stable. On the other hand, the output chords of HMM are fully limited by the chords given in the training examples.

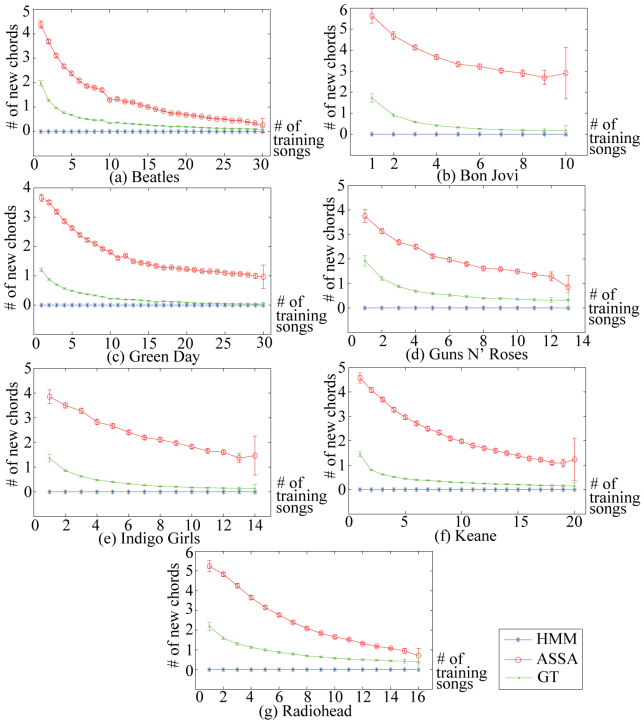

Figure 4. Average number of new chords per song using HMM, ASSA, HA and in the ground truth (GT).

5. CONCLUSIONS AND FUTURE WORK

In this paper we compared three different approaches, a rule-based model, a statistical model, and a hybrid system combining the two, for automatic style-specific harmonization in popular music. We conducted experiments by using songs from several stylistically identifiable pop/rock bands, having the systems generate chords to harmonize given melodies, and compared the generated chords with the original. We observed that the rule-based system generates the most chords within a close range of the original. As the number of training examples increases, the hybrid system reports more chords identical to the original than the other systems. Although the hybrid system has the ability to generate chords that were not present in the training set, it tends to produce too many types of chords for a given song. The HMM-based system, however, produces fewer and fewer chords that are similar to the original as the size of the training set grows. In the future we plan to study different approaches for dividing melodies into melody segments for the harmonization task. We also plan to explore other methods for evaluating system-generated harmonization in a particular style. Besides comparing the generated chords with the original, we will investigate means for measuring the tension and relaxation created in the harmonization.

6. REFERENCES

[1] M. Allan, C. K. I. Williams: "Harmonising Chorales by Probabilistic Inference," *Proceedings of the Neural Information Processing Systems Conference*, Vancouver, 2004.

[2] G. Capuzzo: "Neo-Riemannian Theory and the Analysis of Pop-Rock Music," *Music Theory Spectrum*, vol. 26, no. 2, pp. 177-199, 2004.

[3] C. H. Chuan and E. Chew: "A Hybrid System for Automatic Generation of Style-Specific Accompaniment," *Proceedings of the fourth International Joint Workshop on Computational Creativity*, London, 2007.

[4] K. Ebcioglu, "An Expert System for Harmonizing Four-Part Chorales," *Machine Models of Music*, Cambridge, The MIT Press, 1993.

[5] M. Farbood and B. Schoner: "Analysis and Synthesis of Palestrina-Style Counterpoint Using Markov Chains," *Proceedings of the International Computer Music Conference*, Havana, 2001.

[6] S. Kostka and D. Payne: *Tonal Harmony*, McGraw-Hill, New York, 2003.

[7] H. R. Lee and J. S. Jang, "i-Ring: A System for Humming Transcription and Chord Generation," *Proceedings of the IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan, 2004.

[8] M. Mauch, S. Dixon, M. Casey, C. Harte and B. Fields, "Discovering Chord Idioms through Beatles and Real Book Songs," *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, 2007.

[9] D. Morris, I. Simon and S. Basu, "MySong: Automatic Accompaniment Generation for Vocal Melodies," *Proceedings of Computer-Human Interaction*, Florence, 2008.

[10] S. Phon-Amnuaisuk, A. Tuwson and G. Wiggins, "Evolving Music Harmonization," *Artificial Neural Nets and Genetic Algorithm: Proceedings of Fourth International Conference in Portoroz, Slovenia*, 1999.

[11] D. Temperley: *The Cognition of Basic Musical Structures*, MIT Press, Cambridge, 2004.

[12] D. Temperley and D. Sleator, Harmonic Analyzer, *www.cs.cmu.edu/sleator/harmonic-analysis/*