IMPROVING THE KEY EXTRACTION PERFORMANCE OF A SIMULTANEOUS LOCAL KEY AND CHORD ESTIMATION SYSTEM

Johan Pauwels†*, Jean-Pierre Martens*†*, Marc Leman*‡

†Digital Speech and Signal Processing group (ELIS-DSSP)
‡Institute for Psychoacoustics and Electronic Music (IPEM)
Ghent University, Belgium
johan.pauwels@elis.ugent.be, jean-pierre.martens@elis.ugent.be, marc.leman@ugent.be

ABSTRACT

In this paper, significant improvements of a previously developed key and chord extraction system are proposed. The major improvement is the introduction of a separate acoustic model, designed to verify local key hypotheses. The conducted experimental evaluation shows that the presented system improves the state of the art in local key estimation. Our experimental study further demonstrates that the chord estimation performance is already quite robust, whereas the key estimation performance still happens to be sensitive to a number of factors. In particular, we present figures that illustrate the significant impact of the embedded musicological model and the duration of the processed excerpt on the key estimation accuracy.

Index Terms— Key extraction, chord extraction, music signal processing, music information retrieval

1. INTRODUCTION

The concepts of chords and keys form the basic building blocks of tonal harmony in Western polyphonic music. A chord is defined as a collection of simultaneously sounding notes. It is characterized by a reference note (the root) and tonal distances to this root. A key indicates the properties of a set of notes, played concurrently (forming a chord) as well as sequentially (forming a melody) over a longer period of time. A key is characterized by a tonal center (the tonic) and a distribution of tonal distances in relation to this tonic.

The time frame over which to determine a key is not clearly defined, but it is common to make a distinction between a global key and local keys. The former is based on the musical piece as a whole whereas the latter is defined for successive segments of unequal lengths. Both global key extraction methods [1, 2, 3, 4, 5] and local key extraction methods [6, 7, 8, 9] have been covered in the literature.

From their definitions it follows that the concepts of chords and keys are heavily intertwined with each other. Each played chord raises expectations about the key whereas the key raises expectations about the chords that can be played. This actually was our main motivation (see [10]) for trying to extract them simultaneously and for factorizing the musicological model into maximally independent key and chord transition components. We argue that such a musicological modelling approach first of all complies with the way scholars study harmony, and secondly, that it is bound tot generalize better to unseen data. The latter is actually supported by objective data showing that chord transition models formulated independently of the tonic have a lower perplexity [11].

Although we are not the only ones opting for a simultaneous key and chord extraction (see [5, 6, 7] for example), the majority of systems are designed for either key extraction or chord extraction, and do not explicitly take the relations between both into account. Examples of such approaches to key estimation can be found in [3, 4, 9] and for chord estimation in [12, 13, 14, 15]. Others, like [1, 8], first extract chords and subsequently derive the key from the estimated chords. In one system [2], this cascaded approach is embedded in an iterative system where a first estimation of the chords is refined by the results of a subsequent key analysis step. The reverse also exists, where first the key is determined which is then used as additional feature to extract the chords [16].

The remainder of this paper is constructed as follows. A detailed description of our system is provided in Section 2. During the discussion, the emphasis is on the changes that were made with respect to the baseline described in [10]. In Section 3, we present an experimental study in which we compare our key estimation results with that of other state of the art systems and in which we assess the impact of some factors on these results. In particular, we investigate the impact of the musicological model that guides the estimation process and the influence of the duration of the processed excerpt and its location in the song. We end the paper with some conclusions and ideas for future work.

2. SYSTEM DESCRIPTION

The input audio is first resampled to 8 kHz and converted to mono. The resulting waveform is split into 150 ms long frames with a step size of 20 ms, and for each frame, a chroma profile is calculated which represents the intensity of each of the 12 pitch classes in the frame. The subsequent chroma profiles are then integrated over 11 frames (220 ms), and the integrated profiles are supplied to the back-end. Due to the smoothing, they can be supplied at a rate of one per 220 ms which improves the processing speed.

The back-end traces the most likely state sequence through a finite state machine with 24 * 48 states. Each state represents a distinct key-chord pair. Two key modes — major and minor — are considered for each of 12 possible tonics. Four chord types — major, minor, diminished and augmented — are distinguished for each of 12 possible roots. The search for the best solution is performed by means of an integrated dynamic programming algorithm.

2.1. Chroma extraction

Simply folding a logarithmic frequency spectrum into one octave produces a chroma profile [17] which contains contributions of the harmonics because a note usually contains harmonics of its fundamental frequency. For instance, a third harmonic would add evidence to the chroma a fifth above the fundamental, even though that note has not necessarily been played.

Instead of accounting for harmonics in the templates, like in [15], we chose to deal with this phenomenon in the chroma extraction step. We therefore adopt multiple pitch tracking techniques to resolve partials. A comb filter is used on a peak-picked spectrum to discover harmonic relations between the peaks such that their energy can be assigned to one of the candidate fundamental frequencies once there is enough harmonic evidence to support this hypothesised F0. More details can be found in [18]. Recently, another chroma profile extractor with the same objective has been proposed [6], but here a non-negative least squares algorithm with an idealised note dictionary is utilized to provide an approximate transcription. The advantage of the chosen approach is that it enables the use of binary chord templates in the back-end, and these templates can be directly derived from music theory.

2.2. Probabilistic framework

The back-end implements a unified probabilistic framework for the simultaneous recognition of chords and keys. Its objective is to retrieve the most likely state sequence $\hat{Q}$ for the acoustic observation sequence $\mathbf{X}$. Each state $q_n = (k_n, c_n)$ represents the key and chord combination assigned to a vector $\mathbf{x}_n$. Using Bayes's rule and a first-order Markov assumption, the desired state sequence $\hat{Q}$ follows from

$$\hat{Q} = \arg\max_{Q} \prod_{n=1}^{N} P(q_n|q_{n-1})\, P(\mathbf{x}_n|q_n)$$

with $P(\mathbf{x}_n|q_n)$ representing an acoustic model for estimating the likelihood that chroma profile $\mathbf{x}_n$ is observed when chord c_n is played in a part characterized by key k_n. Because of the scarcity of training data (in comparison with other fields such as speech processing), we did not attempt to train any acoustic models (e.g. Gaussian mixture models that incorporate a lot of free parameters). Instead, we opted for a model that just penalizes the dissimilarities between $\mathbf{x}_n$ and a template vector representing q_n.

In the original system, the acoustic likelihood was simplified to $P(\mathbf{x}_n|c_n)$. We now propose an extension to this and argue that the key and the chord labels can be considered as independent means of testing whether an observation vector complies with a certain state, i.e. $P(\mathbf{x}_n|q_n) = P(\mathbf{x}_n|c_n)P(\mathbf{x}_n|k_n)$. Note that we use the same observations for both key and chord acoustic likelihoods, as opposed to [7] where the observations for the key acoustic likelihoods are integrated over a much longer time than those for the chord likelihoods.

For the chord acoustic model $P(\mathbf{x}_n|c_n)$, the templates consist of binary components: 1 for a chroma that is present in the chord and 0 for one that is not. In [10] two measures for quantifying the similarity between a template and an observation vector were tested: a normalized cosine similarity measure and a product of 12 probabilities for the 12 elements of $\mathbf{x}_n$, derived from two Gaussian models (for templates elements equal to 1 and 0 respectively). The latter gave slightly better results, so we only use that measure as the chord acoustic model.

The key acoustic model $P(\mathbf{x}_n|k_n)$ uses non-binary templates defined by Temperley. These are vectors representing the stability of the 12 pitch classes relative to a given key. They are based on the Krumhansl–Schmuckler profiles, but specifically adjusted for computational key-finding [19]. The measure used here is the normalized cosine similarity between the key templates and the observation vector.

The transition probabilities $P(q_n|q_{n-1})$ are implemented by a compound of three models: a state duration, a key transition and a chord transition model. The state duration model is a simple geometric model that is fully characterized by the chance of staying in the same state. We assume that $P(q_n = q_{n-1}) = P_s$ for all states. The state transition model $P(q_n \neq q_{n-1})$ is decomposed into a key and a chord transition model. Key changes without chord change are prohibited and self-transitions are modelled by the duration model, so the state transition model models effectively $P(c_n \neq c_{n-1})$.

The contribution of the chords to the key transition model is ignored such that we end up with probabilities which are only dependent on the previous key $P(k_n|k_{n-1})$. Our model is based on Lerdahl's regional distance [20, p.68], which expresses numerically the distance between two keys. We make the assumption that keys which are close to each other according to this distance are also likely to appear in sequence and will thus receive a high transition probability. One of the weaknesses in this assumption is that this gives inadequate probabilities for some key changes such as the “gear change” or “one up” which are common in pop music, but not in music of the Common Practice period on which Lerdahl’s theory is based. The distances are converted to probabilities by taking the normalized inverse of the exponential of the distance.

The chord transition model is expressed in terms of relative chords (c'_n, c'_{n-1}) in key k_{n-1}. By doing so, the parallelism between keys differing in tonic but not in mode can be exploited, and key k_{n-1} can be replaced by its mode m_{n-1} in the conditional part of the transition probabilities. This leads to relative chord transition probabilities $P(c'_n|m_{n-1}, c'_{n-1})$. These probabilities can be derived from a set of annotations simply by counting occurrences.

An alternative theoretical model has also been constructed. The transitions that both start and end in a diatonic chord get a probability assigned that is based on Lerdahl’s chord distance within a key [20, p.55], which gives a numerical expression for the distance between two chords in the same key. A similar assumption as for the key transition is made in that a small distance between two chords is converted into a high probability of transition by taking the normalized inverse of the exponential of the distance. Transitions that start or end in a non-diatonic chord (or both) receive a probability that is uniformly distributed. For a deeper explanation of this model, we refer to [10].

For computational reasons, log-probabilities are used and in order to have control over the relative importances of the different sub-models, multiplicative balance parameters α, δ, μ and κ are introduced. Ultimately, a search is performed to find the state sequence which emerges from

$$\hat{Q} = \arg\max_{Q} \sum_{n=1}^{N} \Big[\log P(\mathbf{x_n}|c_n) + \alpha P(\mathbf{x_n}|k_n) + \delta \mathcal{L}_D + \mu \mathcal{L}_M \Big]$$

$$\begin{aligned} \mathcal{L}_D &= \log P_s & (q_n = q_{n-1}) \\ &= \log(1 - P_s) & (q_n \neq q_{n-1}) \\ \mathcal{L}_M &= \big[\kappa \log P(k_n|k_{n-1}) & (c_n \neq c_{n-1}) \\ &\quad + (1-\kappa) \log P(c'_n|m_{n-1}, c'_{n-1})\big] & \\ &= 0 & (c_n = c_{n-1}) \end{aligned}$$

The balancing of the state duration model and the musicological model differs from the balancing approach proposed in [10]. Nevertheless, the baseline system of [10] roughly corresponds to the case of $\alpha = 0$.

3. EXPERIMENTAL RESULTS

In order to assess the performance of our system, we need data with accompanying ground truth labels, as well as an evaluation measure. We have two data sets at our disposal. The first one is the same collection of 142 manually annotated 30 s excerpts of music pieces in a variety of genres and tempi that was used to determine the optimal parameters of the original system, hereafter called the SEMA set. We will also use it here as the development set to optimize the newly introduced parameter α. The second set consists of the 210 songs that were used in the MIREX 2009[1] chord estimation contest. It is composed of full albums by the Beatles (174 songs), Queen (18 songs) and Zweieck (18 songs). The performance is measured as the percentage of time the extracted key or chord equals the annotated key or chord. To avoid a disputable ranking of multiple possible mappings from complex chords to triads, we restrict chord evaluation to segments where one of the basic triads (maj–min–dim–aug, including inversions) was annotated. This leaves us with 62.56% of the data for the SEMA set and 77.44% for the MIREX set. Key extraction performance is measured over the whole data set. Only perfect matches are considered correct, extraction of related keys or chords does not add to the score. We first optimize the free parameters of our system, and then compare it for these optimal settings to other systems that are anticipated to present the current state of the art.

chord trans. model	without $P(\mathbf{x_n}|k_n)$		with $P(\mathbf{x_n}|k_n)$	
	chord	local key	chord	local key
SEMA	73.08	67.44	73.03	70.71
MIREX	72.44	48.97	72.69	58.58
theoretical	72.75	40.17	72.72	51.00

Table 1. Chord and local key performance on SEMA data for 3 different relative chord transition models without and with key acoustic model

3.1. Impact of the key acoustic model

By varying the balance parameter α between 0 and the optimum found in an exhaustive search, we can evaluate the impact of introducing a key acoustic model. We do this for multiple relative chord change models: two trained models (one for each data set) and the aforementioned theoretically derived model. The other free parameters are set to the optimum found in [10].

Tables 1 and 2 show that the key acoustic model causes a substantial improvement of the key estimation performance for both data sets. According to Wilcoxon’s signed rank test, the difference is significant in 5 of the 6 cases ($p < 0.05$ for the MIREX and theoretical chord transition model with the SEMA set and $p < 0.01$ for all chord transition models with the MIREX data). Since improvements on one aspect (key estimation) often results in a degradation on another aspect (chord estimation), we were happy to observe that the chord

[1] http://www.music-ir.org/mirex/wiki/2009:Audio_Chord_Detection

chord trans. model	without $P(\mathbf{x_n}\|k_n)$		with $P(\mathbf{x_n}\|k_n)$	
	chord	local key	chord	local key
SEMA	76.43	59.29	76.38	64.31
MIREX	78.37	73.68	78.40	77.82
theoretical	76.18	59.01	76.11	65.56

Table 2. Chord and local key performance on MIREX data for 3 different relative chord transition models without and with key acoustic model

estimation was not affected[2]. However, we had actually expected an improvement of the chord estimation as well, due to the more reliable key context in which to interpret them. We explain this by arguing that a chord usually fits into multiple keys, and that errors between related keys (adjacent, relative or parallel) will not necessarily induce chord estimation errors. More than 60% of the wrongly estimated keys actually happen to be related keys.

3.2. Impact of the musicological model

Tables 1 and 2 show that a musicological model trained on the test data substantially outperforms a model that is either trained on another dataset or a model relying on dissimilarities derived from music theory: the key estimation accuracy is at least 12% higher and $p < 0.01$ for all cases. Also after the introduction of the key acoustic model, the chord estimation accuracies are not very sensitive to the choice of the musicological model.

3.3. Data set dependency

Although we used parameters that are optimal for the SEMA data set (α found above, the others in [10]), the figures in Table 2 reveal that chord and key estimation is inherently easier on the MIREX set. This does not come as a surprise, since the Beatles are known for their harmony based compositions, while our set was not assembled particularly for chord and key extraction and thus contains a number of more rhythmically oriented songs.

Besides their inherent difference in composition, the two collections differ in another way. The SEMA set consists of 30 seconds excerpts whereas the MIREX set consists of complete songs. In the next section, we describe an experiment we conducted to verify whether this difference could be responsible for part of the observed performance differences.

3.4. Influence of excerpt duration and position

To test the hypothesis that the duration and the position of the processed excerpt could possibly affect the key estimation accuracy, we extracted keys and chords from 30 s and 60 s excerpts of each song of the MIREX set. Furthermore, the test was repeated with excerpts taken from the beginning, the middle and the end of a song. The algorithm settings were not altered. The results of our experiments are summarized in Table 4. They should be compared to the full song results summarized in the right half of Table 2.

The first conclusion we can draw is that neither duration nor position of the excerpt has a noticeable influence on the chord extraction performance.

The local key extraction results confirm the musical intuition that the key is harder to detect if the analysed part is shorter. Nevertheless, the performance for 60 s excerpts is already close to the one for complete songs. The degradation observed for shorter excerpts (30 s) is especially significant in combination with trained musicological models: from around 78 to around 68% when the MIREX model is used and from around 64 to around 57% when the SEMA model is used. That the degradation is larger in combination with the MIREX model may be owed to the fact that excerpts at different positions might exhibit different bigram statistics, meaning that the model trained on full songs is not optimally adapted to the statistics observed in a much shorter excerpt at a particular positions. However, the local key performances for 30 s and 60 s excerpts do not demonstrate any consistent dependency of results on the position of the excerpt: the spread is similar for both excerpt durations, but the tendencies are different.

In combination with a theoretical, distance-based musicological model, the key extraction results do not degrade as much as with the trained models: performance goes from around 66% for full songs to something in between 62 and 67% for 30 s excerpts, even causing an increase in performance for excerpts at the start. We hypothesize that a composition will usually start with rather predictable diatonic chord progressions whereas the less predictable chord progressions only come later, to create tension and ultimately, to assure the continued interest of the listener. If this is true, the theoretical model which clearly favours diatonic chord progressions will comply better with the progressions observed in a song initial excerpt, and the effect will become less apparent if that excerpt becomes longer. Both facts are fully supported by the data. The percentage of diatonic chords can also be directly measured from the annotations. The figures in Table 3 show that there indeed is a higher chance of encountering a diatonic chord in the beginning of a song. The overall ratio of diatonic chords in the MIREX set is 78.27%.

We can now take another look at the results for the two data sets, namely, by comparing the key and chord extraction accuracies obtained with the 30 s excerpts of the MIREX and the SEMA data sets. Since the SEMA excerpts were unfortu-

[2] Because of the differences in our evaluation, the chord performance values for the MIREX data should not be compared to the figures from the MIREX chord extraction contest.

chord	30 seconds excerpts						60 seconds excerpts					
transition	chord			local key			chord			local key		
model	start	middle	end	start	middle	end	start	middle	end	start	middle	end
SEMA	76.21	76.95	75.79	59.03	53.72	59.72	77.02	76.65	76.69	64.07	61.65	61.97
MIREX	77.90	77.95	77.85	70.19	69.01	67.04	78.59	78.16	78.58	73.44	74.21	77.16
theoretical	76.26	76.05	75.17	66.67	61.88	62.21	77.0	75.98	76.25	66.69	65.31	65.84

Table 4. Chord and local key performance on excerpts of MIREX data differing in length and position for 3 different relative chord transition models with key acoustic model

	start	middle	end
30 s	81.33	77.49	79.68
60 s	80.03	77.56	79.45

Table 3. Diatonic chords ratio for excerpts of MIREX data in function of excerpt length and position

nately not taken at a consistent position in the song, we compare each SEMA accuracy with the median of the three corresponding MIREX accuracies. Considering the key extraction results obtained in combination with the best musicological model (the one trained on the test set), we conclude that the differences between the two data sets have completely disappeared. A remarkable difference that persists however is that in combination with the theoretical model, the MIREX data exhibit a much better performance than the SEMA data. Apparently the MIREX chord progressions comply much better to the distance-based theoretical model than the SEMA chord progressions. Considering the chord extraction results, we observe a consistent difference of 3 to 5% in favour of the MIREX results. The latter supports our initial observation of the MIREX set being inherently easier to decode.

3.5. Comparison of key extraction results with the state of the art

The widespread availability of the MIREX data permits us to make a comparison with other algorithms, such as the ones of Mauch & Dixon [6], Rocher et al. [7] and Noland & Sandler [1]. The first two approaches are especially interesting because they also perform simultaneous local key and chord extraction. Mauch & Dixon use a Dynamic Bayesian Network that can model keys and chords as well as metric positions and bass notes. Rocher et al. start from an acoustic model based on template matching. The model is similar to ours, but it employs a multi-scale chroma approach and a back-end which heavily constrains the search space to just a couple of candidate key-chord pairs selected by the acoustic model. Our algorithm on the other hand explores the full search space at all times.

	theirs	ours
Mauch & Dixon	75.58	80.91
Rocher et al.	62.4	80.17 (64.35)
Noland & Sandler	75	86.36

Table 5. Local (M & D, R et al.) or global (N & S) key results in comparison with other systems

Noland & Sandler only estimate the global key of a song by applying an HMM to previously extracted chords. In order to allow for a comparison of that system with ours, we convert our local key sequence to a global key by a simple majority voting. Obviously, more intelligent ways to convert local keys to a global key could have been conceived, but with the just mentioned simple approach, our system achieved the 2nd place in the MIREX 2010 contest for global key extraction. The data set used there was an unseen collection consisting of 1252 excerpts of 30 s each, taken from the beginning of classical music pieces, and synthesized from MIDI. The winning system was a version of the one described in [4], which also produces only global keys. So far, no MIREX competition for local key extraction has been organized.

All three referenced papers report results for different subsets of the MIREX data. Therefore, we have put the results of our system on the same subsets next to the results retrieved from the publications (see Table 5). The results for Mauch & Dixon and Rocher et al. represents local key performance and those for Noland & Sandler global key performance. Since the other algorithms were also tweaked to the MIREX set, we tested our system in combination with the MIREX chord change model. We also worked with a duration model optimized for MIREX. The balance parameters were left unchanged though.

The figures in Table 5 reveal that our system outperforms the state of the art. However, fairness obliges us to say that the algorithm of Rocher et al. was optimized for the Beatles set, but without employing the ground truth labels to set-up a key-chord transition model. They also define a theoretical model instead. If we use our system with our own theoretical model, we obtain the result between brackets in the right column of Table 5. The difference is much smaller then, but still substantial. Noland & Sandler themselves report a result of 91% correct global key estimation when using annotated instead of estimated chord labels. This illustrates the remaining growth potential for audio based key estimation.

4. CONCLUSION AND FUTURE WORK

In this paper we have extended our formerly proposed local key and chord estimation system with a key acoustic model. The experimental evaluation has demonstrated that this extension causes a significant improvement of the local key estimation performance, while the chord estimation remains unchanged. The observed gains are comparable irrespective of the data set.

The integration of specific musicological knowledge in the form of a relative chord change model has proven to have a strong effect on the key estimation results, while it does not alter the chord estimation quality. The latter is also insensitive to the duration and location of excerpts, where the key estimation can vary significantly between different durations and locations. The exact behaviour differs depending on the relative chord model though.

Until now, the musicological model was a simple bigram model. In the future we will try to model the relative chord sequences by means of trigrams, because certain trigrams are known to be excellent indicators of a key. Additionally, a larger context will reduce the perplexity of the chord estimation task. However, we acknowledge the statement of [16] that this does not necessarily lead to a large increase in chord extraction performance.

5. ACKNOWLEDGEMENTS

This work originates from the "GOASEMA" project, funded by the "BOF", Ghent University under contract GOA-1250604. We also thank the Centre for Digital Music at Queen Mary, University of London for providing the MIREX annotations and Matthias Mauch for making the output of his algorithm available online.

6. REFERENCES

[1] K. Noland and M. Sandler, "Influences of signal processing, tone profiles and chord progressions on a model for estimating the musical key from audio," *Computer Music Journal*, vol. 33, no. 1, 2009.

[2] A. Shenoy and Y. Wang, "Key, chord, and rhythm tracking of popular music recordings," *Computer Music Journal*, vol. 29, no. 3, 2005.

[3] S. Pauws, "Musical key extraction from audio," in *Proc. ISMIR*, 2004.

[4] G. Peeters, "Musical key estimation of audio signal based on hidden Markov modelling of chroma vectors," in *Proc. DAFx*, 2006.

[5] K. Lee and M. Slaney, "Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio," *IEEE Trans. Audio, Speech and Language Processing*, vol. 16, no. 2, 2008.

[6] M. Mauch and S. Dixon, "Approximate note transcription for the improved identification of difficult chords," in *Proc. ISMIR*, 2010.

[7] T. Rocher, M. Robine, P. Hanna, and L. Oudre, "Concurrent estimation of chords and keys from audio," in *Proc. ISMIR*, 2010.

[8] H. Papadopoulos and G. Peeters, "Local key estimation based on harmonic and metric structures," in *Proc. DAFx*, 2009.

[9] Ö. İzmirli, "Localized key finding from audio using non-negative matrix factorization for segmentation," in *Proc. ISMIR*, 2007.

[10] J. Pauwels and J.-P. Martens, "Integrating musicological knowledge into a probabilistic system for chord and key extraction," in *Proc. AES 128th Conv.*, 2010.

[11] R. Scholz, E. Vincent, and F. Bimbot, "Robust modeling of musical chord sequences using probabilistic n-grams," in *Proc. ICASSP*, 2009.

[12] C. Harte and M. Sandler, "Automatic chord identification using a quantised chromagram," in *Proc. AES 118th Conv.*, 2005.

[13] J. P. Bello and J. Pickens, "A robust mid-level representation for harmonic content in music signals," in *Proc. ISMIR*, 2005.

[14] A. Sheh and D. Ellis, "Chord segmentation and recognition using EM-trained hidden Markov models," in *Proc. ISMIR*, 2003.

[15] L. Oudre, Y. Grenier, and C. Févotte, "Template-based chord recognition: influence of the chord types," in *Proc. ISMIR*, 2009.

[16] M. Khadkevich and M. Omologo, "Use of hidden Markov models and factored language models for automatic chord recognition," in *Proc. ISMIR*, 2009.

[17] T. Fujishima, "Realtime chord recognition of musical sound: a system using Common Lisp Music," in *Proc. ICMC*, 1999.

[18] M. Varewyck, J. Pauwels, and J.-P. Martens, "A novel chroma representation of polyphonic music based on multiple pitch tracking techniques," in *Proc. ACM Multimedia*, 2008.

[19] D. Temperley, *The cognition of basic musical structures*, MIT Press, 1999.

[20] F. Lerdahl, *Tonal pitch space*, Oxford University Press, New York, 2001.