MIREX 2010: KEY RECOGNITION WITH ZWEIKLANG PROFILES

Andreas Jansson
City University London, School of Informatics
Andreas.Jansson.1@city.ac.uk

Tillman Weyde
City University London, School of Informatics
t.e.weyde@city.ac.uk

ABSTRACT

This extended abstract details a submission to the Music Information Retrieval Evaluation eXchange (MIREX) 2012 for the *Audio Key Detection* task. The system performs key (C, Db, D, E, ...) and mode (major, minor) detection. The system uses a new algorithm called *Zweiklang-Profiling*. A zweiklang is the combination of two pitches, which we determine by detecting the two most prominent chroma values per frame. We calculated frequency profiles of the 145 different Zweiklangs over major and minor modes, similar to the well known pitch class profiles introduced by Krumhansl. The classification algorithm and its implementation is explained briefly in this text.

1. INTRODUCTION

Key recognition from music audio has attracted significant research because the key is important factor in music analysis tasks, such as harmonic and structural analysis, pitch detection, automatic transcription and accompaniment. The *zweiklang profile* algorithm introduced here is an extension of the well known pitch-profile approach introduced by Krumhansl's seminal work [2]. This approach has also been followed and refined by others, like Temperley [5] and Izmirli [1].

2. APPROACH

The *zweiklang profile* extends pitch class profiles to the combinations of two pitch classes, which we call *zweiklang*s, rather than single pitches. We were initially interested in a transition model, similar to Noland and Sandler [3], but the simple *zweiklang* profiles already provide good key classification. For this submission we focused on developing the preprocessing and profile creation.

3. OVERVIEW OF THE ALGORITHM

Zweiklang profiling consists of two phases. First there is a learning phase, where the relative frequencies of the *zweiklangs* relative to the root note are aggregated to key profiles for major and minor. In the application phase the *zweiklang profile* of a piece is calculated. This profile is then correlated with all key profiles, and the profile with the highest correlation determines the answer to return. The profile extraction that is used in both phases consists of audio preprocessing, zweiklang extraction, and profile calculation.

3.1 Audio Preprocessing

The signal is downsampled to 11,025 Hz and sliced into chunks of 4096 samples (370ms). A Fast Fourier Transform s performed on each chunk using a Hanning window. The next step is selecting the peaks in the magnitude spectrum by identifying the local maxima are identified and all other bins are set to 0. For recorded audio we would also apply a quantile filter, that sets the bottom 98 percentiles to 0. We omit this step when applying the program to audio synthesized from MIDI, as in the MIREX. The bins are then folded into one octave to produce a chromagram with 12 bins corresponding to standard semitones. For recorded audio we would also apply a tuning procedure, similar to Rocher at al. [4], that chooses the the slot centres so the energy at the centres is maximised. But again, we omit this step when we expect synthesized input.

3.2 Zweiklang Extraction

From the chromagram the highest value > 0 is selected. If there is none, we report a *Nullklang*. Otherwise the second highest value is selected. If the second peak is exactly one semitone apart from the first it is removed and we choose by the next highest value to avoid artificial values from side lobes. This results in a 2-dimensional vector $\in \mathbb{Z}_{12}^2$, our *zweiklang*. There are 145 zweiklangs: 12×12 different vectors, including those that have only one peak p where both components of the vector are set to p, and the Nullklang.

3.3 Zweiklang Profiles

To extract a *zweiklang profile*, we calculate the relative frequency of the zweiklangs (ignoring Nullklangs) over the relevant piece ore set of pieces and store then in a 144-dimensional vector.

The profiles from the training set are all transposed to root note C. The resulting 2 vectors, one for major and one for minor, are calculated using Laplacian Smoothing before they are used for key recognition.

3.4 Key Recognition

In the recognition phase, the key profile is calculated for the piece of interest without smoothing. It is then compared with each of the 12 major and minor key profiles (obtained by transposition). We use correlation (non-normalised dot-product) as a measure of similarity. We apply a compensation factor to reduce the similarity of minor profiles, which reduced major/minor confusion. The profile with the highest correlation then determines the key.

4. SUMMARY

The Zweiklang-Profile method is a simple method for key detection, that has delivered good result considering its simple structure. We did try to get use a Markov model to include some sequential information, but the results were worse so far. The integration of the Zweiklang approach into more sophisticated models is a task we plan to address in future work.

5. REFERENCES

[1] O. Izmirli: “Template Based Key Finding From Audio”, *Proceedings of the International Computer Music Conference*, 2005.

[2] C. Krumhansl: *Cognitive foundations of musical pitch*, Oxford University Press, Oxford, 1990.

[3] K. Noland and M.B. Sandler: “Key Estimation Using a Hidden Markov Model”, *Proceedings of the International Symposium on Music Information Retrieval*, pp. 121–126, 2006.

[4] T. Rocher, M. Robine, P. Hanna: “Concurrent estimation of chords and keys from audio”, *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010.

[5] D.Temperley: *TheCognitionofBasicMusicalStructures* The MIT Press, Cambridge, 2001.