MIREX-2012 "AUDIO KEY DETECTION" TASK: IRCAMKEYMODE

Geoffroy Peeters

STMS IRCAM-CNRS-UPMC 1, pl. Igor Stravinsky - 75004 Paris - France

ABSTRACT

This extended abstract details a submission to the Music Information Retrieval Evaluation eXchange (MIREX) 2012 for the "Audio Key Detection" task. The system named ircamkeymode performs key (C, Db, D, E, ...) and mode (Major, minor) detection. The system is a simplified version of the systems described in [6] [5]. We briefly summarized it below.

1. OVERVIEW OF THE MODEL

1.1 Chromagram extraction:

The signal is first converted to mono and down-sampled to 11.025 Hz.

At each frame, the DFT of the signal is computed using a Blackman analysis window of length $L = 0.3715$s with a hop-size of $L/2$. After normalization by its maximum value, the amplitude DFT is converted to a Sone scale. The computation of the sone-converted values is similar to the one used in [4]. Thresholding (below 1% of the max value) and peak-picking are then applied.

A 36-bins (3 bins for each semi-tone) chroma representation [7] [1] is then computed. Only frequencies between 100 and 2000 Hz are considered. The shape of the chroma filters is chosen as an hyperbolic tangent with 50% overlap.

Smoothing over time of each of the 36-chroma channels is performed using median filtering.

1.2 Key/Mode templates creation:

We use an approach similar to Gomez [2]: the key profiles are created by extending Krumhansl & Schmukler (Temperley or Diatonic) pitch distribution profile to the polyphonic (several pitches) and audio (several harmonics for each pitch) cases. For each key, we consider the three main triads in this key: the tonic, dominant and subdominant triads (for example in C Major: C-E-G, G-B-D, F-A-C). The chroma vector corresponding to each single note of a specific triad is computed by adding the contribution of its harmonics h. The harmonic h is given a contribution of 0.6^{h-1}. Only the first 4 harmonics are considered. For a specific triad, the chroma vectors corresponding to the three notes are added. Finally for a specific key, the key-chroma vector is computed by adding

the three triad-chroma vectors. Each triad-chroma vector is weighted by the value of the Krumhansl's (Temperley or Diatonic) profile at the position corresponding to the position of the root of the triad in the key (for example 6 for the F-A-C triad in C Major). The result is a 12 dimensions chroma profile vector for each of the 24 keys: $\underline{C_i}$ $i \in [1, 24]$.

1.3 Key/Mode decision:

The most likely key/mode of the track is estimated using an approach similar to Izmirli [3]. The chroma vectors $\underline{c}(t)$ are extracted on a frame basis. At each time t, we estimate the key $\underline{C_i}$ that has the highest correlation (we use the cosine distance) with a cumulated-over-time chroma-vector [1]. We attribute a score to this key proportional to the distance between its correlation value and the correlation value of the second most likely key. This score acts as a reliability coefficient. The final key decision is chosen as the key with the maximum score cumulated over time. Only the first 20 seconds of the tracks are considered.
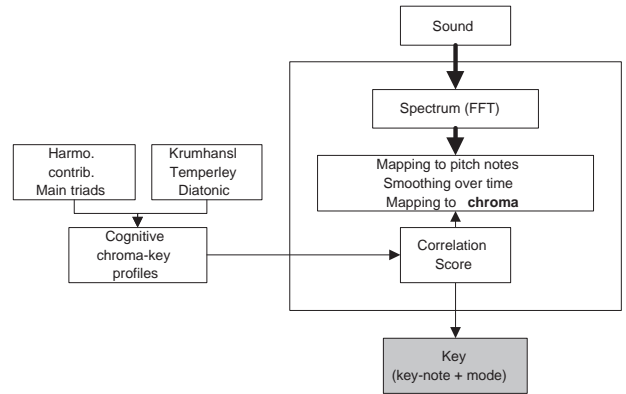
2. FLOWCHART OF THE MODEL

Figure 1. Flowchart of ircamkeymode

3. MIREX-2012 RESULTS AND DISCUSSIONS

TO DO

4. CONCLUSION

TO DO

---

[1] At time t, the cumulated-over-time chroma-vector is computed by averaging the chroma vectors $\underline{c}(\tau)$ since the beginning of the track: $1/t \sum_{\tau=0}^{t} \underline{c}(\tau)$.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] T. Fujishima. Realtime chord recognition of musical sound: a system using common lisp music. In *Proc. of ICMC*, pages 464–467, Bejing, China, 1999.

[2] E. Gomez. Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing, Special Cluster on Computation in Music*, 18(3), 2006.

[3] O. Izmirli. Template based key finding from audio. In *Proc. of ICMC*, pages 211–214, Barcelona, Spain, 2005.

[4] E. Pampalk, S. Dixon, and G. Widmer. Exploring music collections by browsing different views. In *Proc. of ISMIR*, pages 201–208, Baltimore, Maryland, USA, 2003.

[5] G. Peeters. Chroma-based estimation of musical key from audio-signal analysis. In *Proc. of ISMIR*, pages 115–120, Victoria, Canada, 2006.

[6] G. Peeters. Musical key estimation of audio signal based on hmm modeling of chroma vectors. In *Proc. of DAFX*, McGill, Montreal, Canada, 2006.

[7] G. Wakefield. Mathematical representation of joint time-chroma distributions. In *Proc. of SPIE conference on Advanced Signal Processing Algorithms, Architecture and Implementations*, pages 637–645, Denver, Colorado, USA, 1999.