KEYEX: AUDIO KEY EXTRACTION

Steffen Pauws

Philips Research
Prof. Holstlaan 4
5656 AA Eindhoven, the Netherlands
steffen.pauws@philips.com

ABSTRACT

Keyex efficiently computes the *chroma spectrum* from PCM audio data by extracting the presence of all possible musical pitches in the range from A0 (~27.5 Hz) to A6 (1760 Hz) by using *subharmonic summation* in non-overlapping time frames of 100 msecs. These six octaves are collapsed into one octave resulting in a chroma spectrum containing all pitch classes in A-440 equal temperament. This chroma spectrum is used as input in the *key profile matching algorithm* to get the musical key.

Keywords: MIREX contest, audio key extraction.

1 CHROMA SPECTRUM

A chroma/pitch class c_i $(i = 1,\ldots,12)$, represents a set of pitches $\{p_{ik} | k = 1,\ldots,K\}$, where K denotes the number of octaves, that have the same scale position *i*, but that differ in octave position *k*. Thus, telling the presence of a chroma in music requires knowing the presence of all pitches sharing the same chroma. To this end, we convert the music signal into a sequence of spectrum representations $S_1,\ldots,S_T$ of length T. Each spectrum representation models the music signal in a compact form over a short time frame (i.e., 100 msecs). This representation can be an *N*-bin FFT-spectrum but this has shortcomings in frequency resolution and resolving harmonics. Therefore, we use the subharmonic sum spectrum, which has been designed to overcome these problems. For instance, it automatically resolves higher harmonics onto their fundamental (see Section 2).

The problem of telling the presence of a chroma c_i in a spectrum representation S_t has been formulated as maximum likelihood estimation procedure, arriving at the following expression:

$$P(S_t | c_i) \propto \frac{\left| \sum_{k=1}^{K} S_t(p_{ik}) \right|^2}{\left| \sum_{n=1}^{N} S_t(n) - \sum_{k=1}^{K} S_t(p_{ik}) \right|^2},$$

where we assume that p_{ik} corresponds to a bin in the spectrum representation of *N* bins. By computing this expression for each chroma, we arrive at likelihood scores for all chromas in short time frame. This is what we call a chroma spectrum.

2 SUBHARMONIC SUMMATION

Subharmonic summation means adding harmonically compressed amplitude spectrum representations with the following modification:

- spectral content above 5000 Hz is cut off by down-sampling the signal
- the spectral components (i.e., the peaks) are enhanced to cancel out spurious peaks,
- the frequency abscissa is transformed to a logarithmic one by means of cubic interpolation (48 points per octave over 5 octaves), to overcome frequency resolution problems.
- a weighting function is used to model the human auditory sensitivity.

The summation process itself can be best expressed as

$$S(s) = \sum_{m=1}^{M} h^{m-1} W(s + \log m) A(s + \log m)$$

where $s = \log f$ denotes the logarithmic frequency, *M* (e.g., 15) denotes the number of harmonics, *m* is the compression rank, *h* (e.g., 0.84) denotes the decreasing factor, *W(s)* is an arc-tangent function representing the transfer function of the auditory sensitivity filter, and *A(s)* denotes the amplitude FFT spectrum.
Default page size is A4. All text must be in the default two-column format; figures and tables may span two columns where appropriate.

3 EVALUATION

The MIREX 2005 contest compared seven algorithmic approaches to key extraction from music audio files. The dataset consisted of 1,252 audio renditions of MIDI files by Winamp synthesis and Timidity (using Fusion soundfonts) synthesis. In the evaluation, a correct key assignment (as defined by the composer) was given a full point, and incorrect assignments were allocated fractions of a point according to the following table:

Relation	Points
Same	1
Perfect fifth	0.5
Relative major/minor	0.3
Parallel major/minor	0.2

The composite percentage score for keyex was 85.0% (1063.9/1252) with a total run time of about 500 seconds on a standard PC platform.