NOTE ONSET DETECTION USING ONE SEMITONE FILTER-BANK FOR MIREX 2009

Antonio Pertusa
Dpto. Lenguajes y Sistemas Informáticos.
Universidad de Alicante, Spain
pertusa@dlsi.ua.es

José M. Iñesta
Dpto. Lenguajes y Sistemas Informáticos.
Universidad de Alicante, Spain
inesta@dlsi.ua.es

ABSTRACT

The presented onset detection approach is a very simple method described in [1]. An implementation in D2K was already submitted for MIREX 05 [2], yielding a relatively low success rate. However, probably there were some problems in the evaluation, as the mean distance between the detected and actual onsets was too high (about -22 ms, see [3]). Therefore, the system has been reimplemented in C++ and submitted again for evaluation. The methodology and parameters are the same than those described in [1] and [2]. The source code has also been released [1] for research purposes.

1. INTRODUCTION

In a preprocessing stage, the short time Fourier Transform of audio signal is performed and the magnitude spectra are analyzed across a 1/12 octave (one semitone) band-pass filter bank simulated in the frequency domain. Then, the derivatives in time of the filtered values are considered to detect spectral variations related to note onsets.

The motivation of the proposed approach is based on the characteristics of most harmonic tuned instruments. The first harmonics of a tuned sound are close to the frequencies of other pitches in the equal temperament. Other characteristic of musical instruments is that usually most of the energy of the sound is contained in the first harmonics. The one semitone filter bank is composed by a set of triangular filters, which center frequencies coincide with the musical pitches (see Fig. 1).

In the sustain and release stages of a sound, there may be slight variations in the intensity and the frequency of the harmonics. For example, a harmonic peak in the frequency bin k in a given frame can be shifted to the bin $k+1$ in the following frame. In this scenario, direct spectra comparison, like spectral flux, may yield a false positive, as intensity differences are detected. Using this filter bank, the output value of the filter which center is close to k will be similar in both frames, avoiding a false detection.

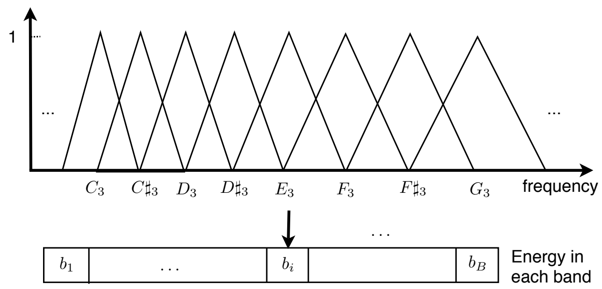

Figure 1. One semitone filter bank.

Therefore, by using one semitone filters, the effects of subtle spectrum variations produced during the sustain and release stages of harmonic sounds are minimized, whereas in the attack stage the filtered amplitudes increase significantly, as most of the energy of the partials is concentrated in the center frequencies of the semitone bands. This way, the system is specially sensitive to frequency variations that are larger than one semitone.

2. METHODOLOGY

For detecting the beginnings of the notes in a musical signal, the method analyzes the spectrum information across one semitone filter bank, computing the band differences in time to obtain a detection function. Peaks in this function are extracted, and those which values are over a constant threshold are considered as onsets.

2.1 Preprocessing

From a digital audio signal, the STFT is computed, providing its magnitude spectrogram. A Hanning window with 92.9 ms length is used with a 46.4 ms hop size. With these values, the temporal resolution achieved is $\Delta t = 46.4$ milliseconds, and the spectral resolution is $\Delta f = 10.77$ Hz.

Using a 1/12 octave filter bank, the band corresponding to G$\sharp_0$ has a center frequency of 51.91 Hz, and the fundamental frequency of the next pitch, A$_0$, is 55.00 Hz, therefore this spectral resolution is not enough to build the lower filters. Zero padding has been used to get more spectral bins, appending $3 \cdot 2048$ zero samples to each frame

[1] Source code can be downloaded from http://grfia.dlsi.ua.es/cm/worklines/pertusa/onset/pertusa_onset.tgz

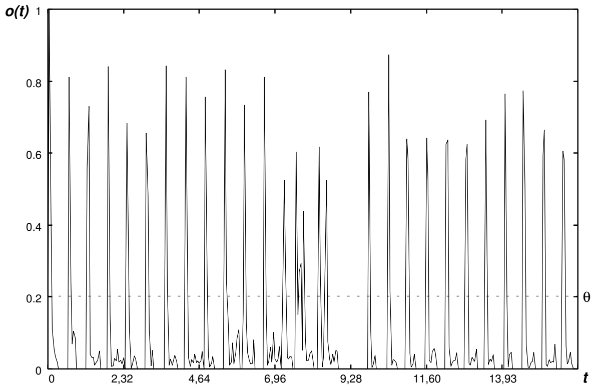

Figure 2. Example of the onset detection function $o[t]$ for a piano melody, RWC-MDB-C-2001 No. 27 from [5].

before computing the STFT. With zero padding, the frequency spacing is $\Delta f = 10.77/4 = 2.69$ Hz.

At each frame, the spectrum is apportioned among a one semitone filter bank to produce the corresponding filtered values. The filter bank comprises from 52 Hz (pitch $G\sharp_0$) to 10,600 Hz (pitch F_8), almost eight octaves. This way, $B = 94$ filters are used, which center frequencies correspond to the fundamental frequencies of the 94 notes in that range. The filtered output at each frame is a vector $\mathbf{b}$ with B elements.

$$\mathbf{b} = \{b_1, b_2, \ldots, b_i, \ldots, b_B\} \tag{1}$$

Each value b_i is obtained from the frequency response H_i of the corresponding triangular filter i with the spectrum:

$$b_i = \sqrt{\sum_{k=0}^{K-1} (|X[k]| \cdot |H_i[k]|)^2} \tag{2}$$

2.2 Onset detection function

Like in other onset detection algorithms, a first order derivative function is used to select potential onset candidates. In the presented approach, the derivative $c[t]$ is computed for each filter i.

$$c_i[t] = \frac{\mathrm{d}}{\mathrm{d}t} b_i[t] \tag{3}$$

These values must be combined to yield the onsets in the overall signal. In order to detect only the beginnings of the events, the positive first order derivatives of all the bands are summed at each time, whereas negative derivatives are discarded:

$$a[t] = \sum_{i=1}^{B} \max\{0, c_i[t]\}. \tag{4}$$

To normalize the onset detection function, the overall energy $s[t]$ is also computed (note that $a[t] < s[t]$):

$$s[t] = \sum_{i=1}^{B} b_i[t] \tag{5}$$

Participant	Params	F-m %	Pr %	Re %	Runtime
AR2	0.43	79.60	85.00	79.19	00:04
AR1	0.34	79.00	81.32	83.30	00:03
AR5	0.34	78.31	80.56	81.88	00:03
PI	0.25	76.79	79. 99	77.50	00:01
AR4	0.13	76.48	86.39	73.62	00:02
AR3	0.28	76.10	85.96	73.15	00:02
TZC1	N/A	74.43	75.67	76.97	01:57
TZC2	N/A	73.38	78.28	74.58	01:57
TZC3	N/A	68.63	79.61	68.97	01:57
TZC5	N/A	68.23	62.88	83.69	01:57
TZC4	N/A	67.94	78.98	68.91	01:57
GT	N/A	59.54	67.01	59.91	00:01

Table 1. Overall MIREX 2009 onset detection results ordered by average F-measure. Runtime is measured in hh:mm.

Participant	Params	F-m %	Pr %	Re %
AR5	0.34	91.56	92.07	92.02
AR1	0.34	91.54	91.70	92.31
PI	0.2	90.36	93.58	88.27
AR2	0.52	89.20	93.02	87.79
AR4	0.43	87.26	98.51	80.21
AR3	0.49	87.11	96.12	80.97
TZC1	N/A	83.12	85.94	83.11
TZC2	N/A	83.12	85.94	83.11
TZC3	N/A	74.17	89.61	70.73
TZC4	N/A	73.77	89.23	70.32
TZC5	N/A	68.66	61.33	90.01
GT	N/A	67.41	71.91	66.43

Table 2. MIREX 2009 poly-pitched results ordered by average F-measure.

The sum of the positive derivatives $a[t]$ is divided by the sum of the filtered values $s[t]$ to get a normalized relative difference. Therefore, the onset detection function $o[t] \in [0, 1]$ is:

$$o[t] = \frac{a[t]}{s[t]} = \frac{\sum_{i=1}^{B} \max\{0, c_i[t]\}}{\sum_{i=1}^{B} b_i[t]} \tag{6}$$

2.3 Peak detection and thresholding

The last stage is to select the onsets from $o[t]$. Peaks at time t are identified in the onset detection function when $o[t-1] < o[t] > o[t+1]$, and those peaks over a fixed threshold $o[t] > \theta$ are considered as onsets. Two consecutive peaks can not be detected, therefore the minimum temporal distance between two onsets is $2\Delta t = 92.8$ ms. A silence threshold μ is also introduced to avoid false positive onsets in quiet regions, in such a way than if $s[t] < \mu$, then $o[t] = 0$.

The silence gate μ is only useful when silences occur, or when the considered frame contains very low energy, therefore it is not a critical parameter. The precision/recall deviation can be controlled through the threshold θ.

Fig. 2 shows an example of the onset detection function $o[t]$ for a piano piece, where all the peaks over the threshold θ were correctly detected onsets.

3. RESULTS

The detailed evaluation results can be found in [4]. The overall results are shown in Tab. 1. The proposed method yielded a competitive F-measure with a very low computational cost. As it was primarily designed for detecting spectral changes of more than one semitone, the poly-pitched results (see Tab. 2) are of special interest for this method and, for these kind of sounds, the F-measure was close to the best. However, the proposed approach also seems to yield good results with unpitched sounds, as it obtained the highest F-measure in solo-drum excerpts.

4. ACKNOWLEDGEMENTS

This project is supported by the Spanish CICyT PROSEMUS project (TIN2006-14932-02) and the Consolider Ingenio 2010 program (MIPRCV project, CSD2007-00018).

5. REFERENCES

[1] A. Pertusa, A. Klapuri and J. M. Iñesta: "Recognition of note onsets in digital music using semitone bands," *Lecture Notes in Computer Science*, Vol. 3773, pp. 869–879, 2005

[2] A. Pertusa, A. Klapuri and J. M. Iñesta: "Note detection using semitone bands," In [3], 2005

[3] MIREX 2005, note onset detection contest: `http://www.music-ir.org/evaluation/mirex-results/audio-onset/`

[4] MIREX 2009, note onset detection contest: `http://www.music-ir.org/mirex/2009/index.php/Audio_Onset_Detection_Results`

[5] M. Goto, H. Hashiguchi, T. Nishimura and R. Oka: "RWC Music Database: Popular, Classical, and Jazz Music Databases," In *Proc. of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, pp. 287–288