

# Note onset detection using semitone bands

**Antonio Pertusa**

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

**Anssi Klapuri**

Signal Processing Laboratory  
University of Technology, Finland  
klap@cs.tut.fi

**José M. Iñesta**

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

## ABSTRACT

A simple note onset detection system for music is presented in this work. It is not our aim to use a perceptually motivated approach. Instead, a musically motivated filter-bank is utilized. In the well-tempered scale, the one used in western music, the first harmonics of the tuned instrument notes are close to the center frequencies of the 1/12 octave (one semitone) bands. In most instruments these first harmonics are those with the highest amplitudes. In music, notes are separated by semitones, so the system tries to find semitone variations, which usually correspond to note onsets.

From a digital audio file, a short-time Fourier transform (STFT) is computed, providing its spectrogram. The STFT is calculated using a Hanning window with an overlapping percentage of 50% in order to retain the information at the frame boundaries. With these parameters, the time resolution is  $\Delta t = 46.4$  milliseconds.

The spectral bins obtained are analyzed into 79 semitone bands in a logarithmic scale, ranging from 150 Hz to 22050 Hz. The center frequencies of the spectral bands correspond to the fundamental frequencies of the musical pitches.

To analyse the bins into the bands, a Root Mean Square computation is performed to emphasize the highest spectrum values. A simple equation to get each band value  $b_k(t)$  at time  $t$  can be used:

$$b_k(t) = \sqrt{\sum_{j=1}^{W_k} (X(j, t)w_{kj})^2}, \quad (1)$$

being  $\{w_{kj}\}_{j=1}^{W_k}$  the triangular window values for each band,  $W_k$  the size of the  $k$ -th window and  $X$  the set of spectrum bins corresponding to that window at time  $t$ , with  $j$  indexing the frequency bin.

At a given time, a first order derivative function  $c(t)$  is computed for each band  $k$ .

$$c_k(t) = \frac{d}{dt}b_k(t) \quad (2)$$

In order to detect only the beginnings of the notes, the positive first order derivatives of all the bands are summed at each time and the negative derivatives are not considered.

$$a(t) = \sum_{k=1}^B \max\{0, c_k(t)\}. \quad (3)$$

To obtain a normalized onset detection function, the overall sum of the band amplitudes  $s(t)$  is also computed.

$$s(t) = \sum_{k=1}^B b_k(t) \quad (4)$$

Finally, the sum of the positive derivatives is divided by the overall sum of the band amplitudes at a given time.

$$o(t) = \frac{a(t)}{s(t)}. \quad (5)$$

A silence threshold  $\mu = 79.0$  is applied to the sum of the band amplitudes, in order to avoid false positives when nothing is sounding. The onset detection function peaks that are over a fixed low level threshold  $\theta = 0.18$  are finally considered as onsets.

The system was evaluated for MIREX, and the overall success rate obtained (using the average F-measure) was 58.92%. This algorithm is intended to work with tuned musical instruments, so it is not adequate for voice (11.12%) and it do not seem to be the best option for drums (77.22%). The experiments were also done using complex melodies (50.16%), poly-pitched (59.37%), bars and bells (60.22%), brass (54.41%), plucked strings (67.74%), wind (25.58%) and sustained strings (39.79%). The results are lower than most of the algorithms compared, but it is a first basic prototype and we hope to make several improvements to get better results.

## Acknowledgements

This work has been funded by the Spanish CICYT project TIRIG with code TIC2003-08496-C04, partially supported by European Union-FEDER funds, and the Generalitat Valenciana project with code GV04B-541. Thanks to Jasón Box for migrating the onset detector C++ code into D2K.