Detection of distinct sound events in acoustic signals (onset detection)

Technical report for the course *Psykoakustiikan kesäseminaari 1999*
Anssi Klapuri, klap@cs.tut.fi

Introduction

Onset detection plays an important role in the computational segmentation and analysis of acoustic signals. It facilitates cut-and-paste operations and editing of audio recordings. The onset information may also be used in audio/video synchronization and timing, or passed for further analysis and recognition for example in an acoustic supervision system.

The term *onset detection* is used to refer to the detection of the beginnings of discrete events in acoustic signals. A percept of an onset is caused by a noticeable change in the intensity, pitch or timbre of the sound. A fundamental problem in the design of an onset detection system is distinguishing genuine onsets from gradual changes and modulations that take place during the ringing of a sound. This is also the reason why robust one-by-one detection of onsets has proved to be very hard to attain without significantly limiting the set of application signals.

A system was designed, which is able to detect the perceptual onsets of sounds in acoustic signals. The system is general in regard to the sounds involved and was found to be robust for different kinds of background interference. The system is also able to determine the beginnings of sounds that exhibit onset imperfections, i.e., the amplitude envelope of which does not rise monotonically. Calculation is done at separate frequency bands, and a psychoacoustic model of intensity coding is used to combine the results from the bands. The performance of the system was validated by applying it to the detection of onsets in musical signals that ranged from rock to classical and big band recordings.

Code available

The system consists of Matlab and C-files, all of which can be found in the directory
    /share/arg/klap/jussi/onsets
C-files must be compiled from *.c into *.mexsol files before they can be called from Matlab. This can be done using “mex file.c” command from Matlab command line for each C-file (see “help mex”). Readily compiled versions for SUN Sparc processor can be found in the mentioned directory.

Usage of the system:
    [times, louds] = onsetsAK(signal, srate);
Input consists of the time domain signal and its sample rate. Output consists of the times (samples) and loudnesses (dB) of the beginnings of each distict acoustic event in the input signal. In order to get accurate results, the input signal should be at least couple of seconds long (i.e., between 3-30 seconds).

More information on parameters etc.: “help onsetsAK”.

System overview

The earliest onset detection systems typically tried to process the amplitude envelope of a signal as a whole. Since this was not very effective, later proposals have evolved towards bandwise processing. Scheirer was the first to clearly point out the fact that an onset detection algorithm should follow the human auditory system by treating frequency bands separately and then combining results in the end [Scheirer].

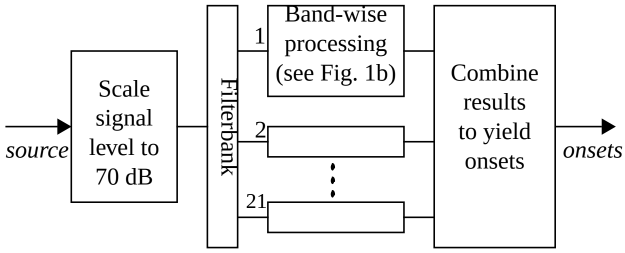

Figure 1a. System overview.

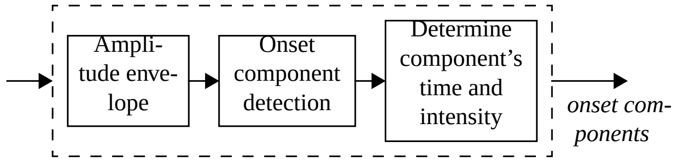

Figure 1b. Processing at each frequency band.

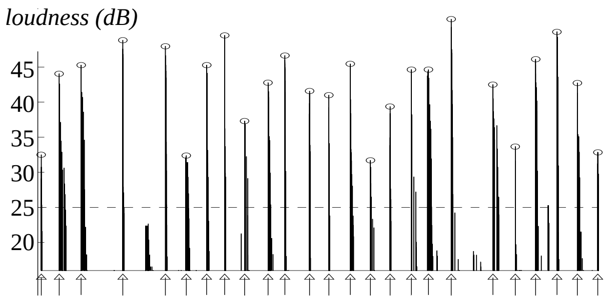

Figure 2. The loudness of onsets as a function of their time. The genuine onsets can now be quite easily discerned by their loudness.

The overview of our onset detection system is presented in Figure 1. It utilizes the band-wise processing principle as motivated above. First, the overall loudness of the signal is normalized to 70 dB level using the model of loudness as proposed by Moore et al. [Moore]. Then a filter-bank divides the signal into 8 non-overlapping bands. At each band, we detect *onset components* and determine their time and intensity. In final phase, the onset components are combined to yield onsets. The system is described in more detail in [Klapuri].

In the final phase, beginnings of sound events are distinguished from other acoustic activity by a simple peak picking operation. This operation, as illustrated in Figure 2, is based on a simple peak picking, which accepts onsets above an absolute threshold as true ones.

Results

Musical signals

The presented system was verified by testing its performance in detecting onsets in musical signals. The signals were selected to comprise a large variation of sound sources (musical instruments) and a wide dynamic and pitch range. Signals both with and without drums were included. Another goal was to include representative excerpts from different musical genres, ranging from jazz and rock to classical and big band music.

Some example results for musical signals are given in Figure 3. It should be noticed, that the rhythmic regularity of music was *not* utilized in any manner, but beginnings of each event were detected one-by-one. In results, quantized time can be observed (characteristic to music), and the distribution of the onsets agrees with the rhythmic percept of the acoustic signals.

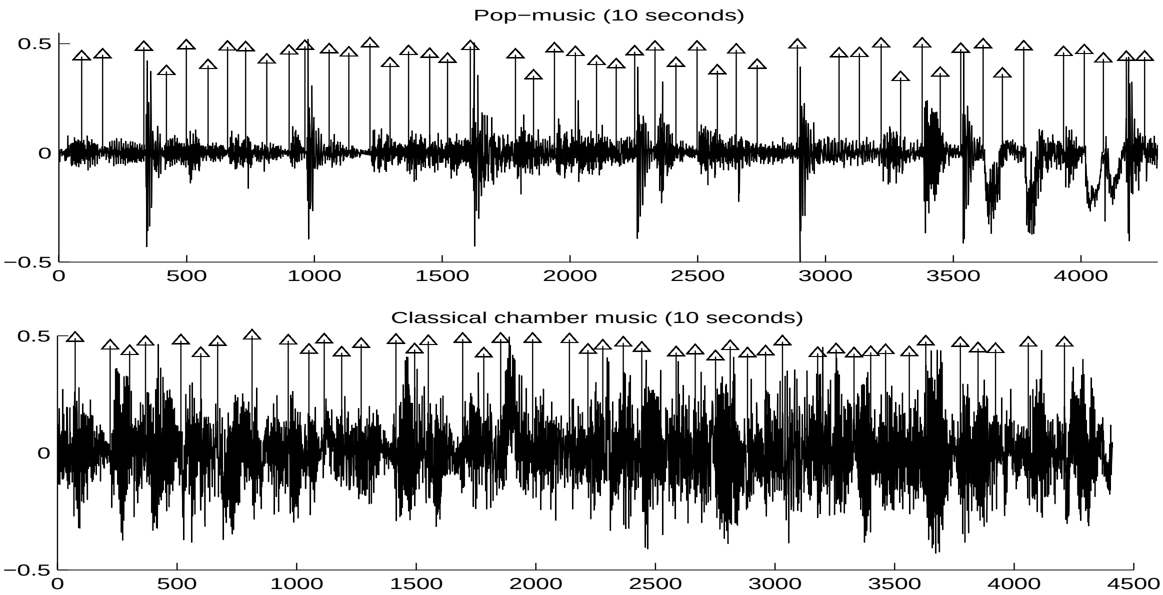

Figure 3. Results for two musical signals. Rhythmic regularity of music was not utilized in any manner, but onsets were detected one-by-one. Distribution of the onsets agrees with the rhythmic percept of the acoustic signals.

Speech

The problem of detecting distinct acoustic events is not well defined for continuous speech signals. The question is: should we detect single phonems, syllables, words, or something else.

However, the system gives in a way “intuitive” results for speech signals. Detected onsets for a single speaker (female) sentence are presented in Figure 4. It seems that the system places onsets at the positions of strong consonant phonems..

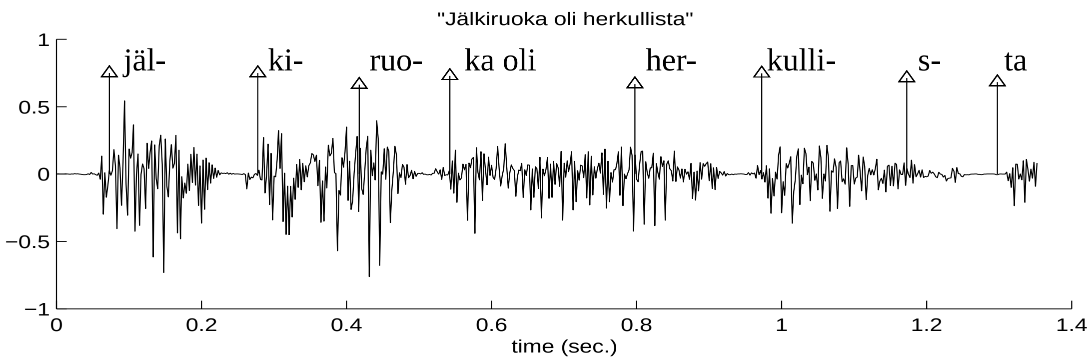

Figure 4. Detected onsets for a single speaker (female) sentence. Onsets are detected in the positions of strong consonant phonems.

References

Scheirer E. “Tempo and Beat Analysis of Acoustic Musical Signals”. Machine Listening Group, MIT Media Laboratory, 1996.

Moore B., Glasberg B., Baer T. “A Model for the Prediction of Thresholds, Loudness, and Partial Loudness”. J. Audio Eng. Soc., Vol. 45, No. 4, pp. 224–240. April 1997.

Klapuri A. “Sound onset detection by applying psychoacoustic knowledge”. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1999.