ONSET DETECTION BY MEANS OF TRANSIENT PEAK CLASSIFICATION

A. Röbel
IRCAM-CNRS-UPMC –STMS
1, pl Igor-Stravinsky
75004 Paris, France
roebel(at)ircam.fr

ABSTRACT

The extended abstract describes an onset detection algorithm that is based on a classification of spectral peaks into transient and non-transient peaks and a statistical model of the classification results to prevent detection of random transient peaks due to noise. Compared to the version used for MIREX 2010 the algorithm presented here differs with respect to the treatment of soft harmoinc onsets.

1. INTRODUCTION

In the following article we are going to describe a transient detection algorithm that has been developed for a special application, the detection of transients to prevent transformation artifacts in phase vocoder based (real time) signal transformations [8, 9]. This application requires a number of special features that distinguishes the proposed algorithm from general case onset detection algorithms: The detection delay should be as short as possible, frequency resolution should be high such that it becomes possible to distinguish spectral peaks that are related to transient and non transient signal components, for proper phase reinitialization the onset detector needs to provide a precise estimate of the location of the steepest ascend of the energy of the attack. In contrast to this constraints the application does not require the detection of soft onsets, where a soft onset is characterized by time constants equal to or above the length of the analysis window. This is due to the fact that such onsets are sufficiently well treated by the standard phase vocoder algorithm. False positive detections are not very problematic as long as they appear in noisy time frequency regions. A major distinction is that a single onset may be (and very often is) composed of multiple transient parts, related either to a slight desynchronization of polyphonic onsets or due to sound made during the preparation of the sound (gliding fingers on a string). While these desynchronized transients are generally not considered as independent onsets they nevertheless constitute transients which should be detected for the intended application.

The evaluation of the transient detection algorithm for onset detection has been evaluated repeatedly in the MIREX evaluation campaigns 2005, [5], 2006 [6], 2007 [7], and 2009 [**?**] and it has shown very good performance many of these evaluations. The analysis of the performance with respect to onset and instrument classes shows that one of the problems of all algorithms is related to the detection of onsets of pitched instruments. Accordingly we have worked on this problem and present here the results of the work. An investigation of the resultts of the previous algorithms has shown that in some situations indivdiual onsets are fragmented resulting in a reduced attack strength which will in turn have a bad impact on our transient filtering procedure.

2. FUNDAMENTAL STRATEGY

There exist many approaches to detect attack transients. For a number of current approaches see [2–4, 12] as well as all algorithms presentd in the MIREX campaigns mentioned above. Most of the known algorithms define an onset detection function that is evaluated in different frequency bands. Here we use a similar approach using as detetcion function a statistical measure related to the time offset (time reassignment) [1] of individuel spectral peaks in the standard DFT spectrum. Using a simple threshold for the time reassignemnt we classifiy spectral peaks into transient and non transient peaks [8, 9] and use as detection function the change in the transient peak probability in the different spectral bands. The advantage of the implicit peak classification is the fact that for each detected transient we have a precise measure of the time frequency location of the related transient.

The basic idea of the proposed transient detection scheme is straightforward. A peak is detected as potentially transient whenever the center of gravity (COG) of the time domain energy of the signal related to this peak is at the far right side of the center of the signal window. Note, that it can be shown [10] that the COG of the energy of the time signal and the normalized energy slope are two quantities with qualitatively similar evolution and, therefore, the use of the COG of the energy for transient detection instead of the energy evolution appears to be of minor importance.

3. FROM TRANSIENT PEAKS TO ONSETS

Unfortunately not every spectral peak detected as transient indicates the existence of an onset. Further inspection reveals that spectral peaks related to noise signals quite often have a COG far of the center of the window. In contrast to spectral peaks related to signal onsets these false transient peaks in noise are not synchronized in time with respect to each other. This synchronization of a sufficient number of transient peaks is the final means to avoid detection of noise peaks as onsets.

To keep this abstract brief we will not describe the details of the statistical model, and we refer to the description of the first mirex evaluations for further details [10, 11].

4. PITCHED TRANSIENTS

The onset detection algorithm that is presented here is based on the detection of multiple synchronous events in the detection bands. The bands that have been used in the previous versions of the algorithm where always covering continuous frequency regions. In a polyphonic setting this band organization is not apprpriate for soft pitched onsets, because these soft onsets will not be detected by the rather high threshold of the COG that can be used for quick onsets. Cutting the frequency band into harmoinc subbands did not bring the expected benefit [**?**]. In the present implementation we therefore propose a different approach that does not try to address the harmonic band structure but the softness of the attacks. The idea is to lower the COG threshold for harmonically related energy displacement, that is whenever we observe a COG above a certain threshold (below the COG threshold used for non harmoinc transients) then these peaks are counted as transient peaks in the statistcial transient probablity model [10, 11].

This approach allows us to lower the sensitvity of the transient detector whenever the transients appear in a harmonic structure. The parameters to select are related to the number of harmonic soft transients that are required to trigger the peak conversion.

5. PARAMETER ADAPTATION

Compared to previous MIREX contributions w significantly extended our training database. We hope that the selected parameters will prove to more appropriate when confronted with different real world signals used in the MIREX evaluation. One of the least theoretically founded parameters - the minimum transient amplitude threshold - has been removed from some of the examples to see whether the algorithm can be configured to work robustly without this parameter that does not provide any relevant

6. DIFFERENCES IN THE 5 SUBMITTED ONSET DETETCTION ALGORITHMS

The submissions use 5 different configurations that can be produced with the 2 algorithmic extensions:

- *roebel_onset_ldb_noharm* baseline version, does not make use of harmonic onsets extensions,
- *roebel_onset_ldb_noharm_ma* baseline version, does not make use of harmonic onsets extensions but uses a weak filter on signal amplitudes
- *roebel_onset_ldb_harm* transient detection using harmonic onset detection
- *roebel_onset_ldb_harm_ma* transient detection using harmonic onset detection and a weak filter on signal amplitudes.
- *roebel_onset_ldb_harm_ma_sw* transient detection using harmonic onset detection and a weak filter on signal amplitudes with short analysis window.

7. REFERENCES

[1] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. on Signal Processing*, 43(5):1068–1089, 1995.

[2] J. Bonada. Automatic technique in frequency domain for near-lossless time-scale modification of audio. In *Proc. of the Int. Computer Music Conference (ICMC)*, pages 396–399, 2000.

[3] C. Duxbury, M. Davies, and M. Sandler. Improved time-scaling of musical audio using phase locking at transients. In *112th AES Convention*, 2002. Convention Paper 5530.

[4] P. Masri and A. Bateman. Improved modelling of attack transients in music analysis-resynthesis. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 100–103, 1996.

[5] Mirex audio onset detection evaluation results. `http://www.music-ir.org/evaluation/mirex-results/audio-onset/index.html`, September 2005. ISMIR 2005, London, Great Britain.

[6] Mirex audio onset detection evaluation results. `http://www.music-ir.org/mirex2006/index.php/Audio_Onset_Detection_Results`, October 2006. ISMIR 2006, Victoria, Canada.

[7] Mirex audio onset detection evaluation results. `http://www.music-ir.org/mirex/2007/index.php/Audio_Onset_Detection_Results`, September 2007. ISMIR 2007, Vienna, Austria.

[8] A. Röbel. A new approach to transient processing in the phase vocoder. In *Proc. of the 6th Int. Conf. on Digital Audio Effects (DAFx03)*, pages 344–349, 2003.

[9] A. Röbel. Transient detection and preservation in the phase vocoder. In *Proc. Int. Computer Music Conference (ICMC)*, pages 247–250, 2003.

[10] A. Röbel. Onset detection in polyphonic signals by means of transient peak classification. http://www.music-ir.org/evaluation/mirex-results/articles/onset/roebel.pdf, September 2005. ISMIR 2005, London, Great Britain.

[11] A. Röbel. Onset detection in polyphonic signals by means of transient peak classification. http://www.music-ir.org/evaluation/MIREX/2006_abstracts/OD_roebel.pdf, October 2006. ISMIR 2005, London, Great Britain.

[12] X. Rodet and F. Jaillet. Detection and modeling of fast attack transients. In *Proc. Int. Computer Music Conference (ICMC)*, pages 30–33, 2001.