ONSET DETECTION BY MEANS OF TRANSIENT PEAK CLASSIFICATION IN HARMONIC BANDS

A. Röbel
IRCAM-CNRS –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 2007 this algorithm focuses on the improvment of the detection of onsets of pitched notes.

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 [1, 2]. 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, [3], 2006 [4] and 2007 [5] and it has shown very good performance at least in the last 2 evaluations. The analysis of the performance with respect to onset and instrument classes shows clearly that all algorithms are comparatively weak when it comes to the detctenio of onsets of picthed instruments. Accordingly we have workd on this problem and present here the results of the work.

2. FUNDAMENTAL STRATEGY

There exist many approaches to detect attack transients. For a number of current approaches see [6–9] 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) [10] 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 [1, 2] 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 [11] 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 [11, 12].

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 versionas of the algorithm where always covering continuous frequency regions. In a polyphonic setting this band organization is a drawback for soft pitched onsets, because these onsets will be characterized by synchronous partials distributed over a large and non continuous frequency band. This systematic problem can be countered easily by means of allowing non continuous observation bands. In the present case we consider observation bands that are formed by a collection of bands with harmonically related center frequencies and a common bandwidth additionnally to the continuous bands that have been used before. The level of confidence of the change in transient peak probability that is required for the detection of a transient event in the non continuous bands can be selected independently of the confidence that is required for the continuos bands. This should allow us to configure the algorithm for different types of sound signals.

5. DIFFERENCES IN THE 5 SUBMITTED ONSET DETETCTION ALGORITHMS

The submissions mainly differ with respect to the selected parameter sets. The parameters have been optimized by means of a genetic algorithm using different sound data bases as follows. The algorithms marked as 12_nhd and 16_nhd have been trained on the same data sets that I had used for the MIREX submissions 2005-2007. The data sets differ only due to minor corrections in the onset labels. The algorithms marked as 7_hd, 10_hd and 19_hdc have been trained on an extended data set that includes some new sounds with purely tonal instruments. These additional sounds have been generated with a midi synthesizer according to [13]. These 2 parameter sets use a longer analysis window and therefore, they should be better suited for polyphonic sound signals. The algorithm used in 19_hdc is slightly different from the others in that it uses a weighting scheme to improve detection of onsets for repeated notes. It is work in progress and may be buggy.

Besides the fact that the algoroithm has been chnaged we updates as well the determination of the final transient position. Compared to the strategy used in last years algorithms where the transient marker hadd been placed at the end of the transient region (sound segment containing start and end of transient events) we now moved the transient position into the center of the transient region.

6. DISCUSSION OF THE RESULTS

The adaptation of the detection parameters has been done on two different training datbases. On one hand the database that we had used to adapt parameters for the previous MIREX campaigns and on the other hand an enlarged database of reference sound exmples that added a new class of soft harmonic onsets to the training database. In total for the evaluation on our training databases we did improve the detection results with the new settings. This was true for the use of the old (fast attack transient only) or the the new training database (enhanced with soft and harmonic onsets).

In the mirex competition, however, the new settings did not improve the results compared when compared to last year. The algorithms that had been trained on the extended database do obtain better results. The algorithms that had been trained on the old database (used for MIREX2007) did show worse performance.

2 explanations are possible: either the new placement of the transient markers, that are now in the center of the transient region, is less wel l adapted to the marker placement used for the MIREX data, or the new hamronic band placement for transient markers creates too many false alarms.

7. REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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.

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

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] C. Yeh, N. Bogaards, and A. Röbel. Synthesized polyphonic music database with verifiable ground truth for multiple f0 estimation. In *Proc. of the 8th Int. Conf. Music Information Retrieval (ISMIR 07)*, 2007.