ONSET DETECTION IN POLYPHONIC SIGNALS BY MEANS OF TRANSIENT PEAK CLASSIFICATION - VERSION 3

A. Röbel
IRCAM-CNRS –STMS
1, pl Igor-Stravinsky
75004 Paris, France

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 last years mirex contribution the algorithm has not been modified, but bug fixes and a better parameter optimization procedure should lead to improved performance.

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 [5, 6]. 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 and music segmentation tasks has revealed that the detection results are comparable with existing algorithms for onset detection or signal segmentation tasks [7]. Therefore, it is now the major means for signal segmentation and onset detection in IRCAM's AudioSculpt application [1]. Since MIREX 2006 a number of improvements of the parameter adaptation procedure have been realized and a number of bugs in the algorithm have been fixed. We are interested to see whether the improvements that we realized on our own database will be observed as well in the MIREX evaluation.

In the following article we will briefly describe the algorithm.

2 FUNDAMENTAL STRATEGY

There exist many approaches to detect attack transients. For a number of current approaches see [2, 4, 3, 8]. In contrast to the evaluation of energy evolution in integral frequency bands, a criterion that most of the approaches are relying on, the following article proposes a two stage strategy which first classifies the spectral peaks in a standard DFT spectrum into peaks that potentially may be part of an attack transient and those that are not. Based on this classification a statistical model of background transient peak activity is employed to detect transient events. The advantage of this two stage approach is that the transient components of the signal are classified with rather high frequency resolution, allowing a precise distinction between transient and non transient signal components.

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 [7] 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, these details will be provided in the final article.

4 DIFFERENCES IN THE 3 SUBMISSIONS

The 4 submissions differ slightly with respect to the internal parameters. All are based on the submission 3 of MIREX 2005 which provided overall the best results. The parameters have been optimized by means of a genetic algorithm using our own reference data base.

5 REFERENCES

[1] N. Bogaards, A. Röbel, and X. Rodet. Sound analysis and processing with AudioSculpt 2. In *Proc. Int. Computer Music Conference (ICMC)*, 2004.

[2] 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.

[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] 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.

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

[7] Axel Röbel. Onset detection in polyphonic signals by means of transient peak classification. In *MIREX Online Proceedings (ISMIR 2005)*, London, Great Britain, September 2005. avail. at `http://www.music-ir.org/evaluation/mirex-results/articles/onset/roebel.pdf`.

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