ADAPTIVE WHITENING PREPROCESSING APPLIED TO ONSET DETECTORS, MIREX 2007

Dan Stowell, Mark Plumbley
Centre for Digital Music
Queen Mary, University of London

ABSTRACT

Adaptive whitening is a preprocessing technique for STFT-based onset detectors, which can improve the performance of onset detectors by compensating for spectral rolloff and dynamic variability [7]. In this MIREX submission we applied adaptive whitening to a selection of STFT-based onset detection functions (ODFs) from recent literature. Our code is oriented primarily towards use in real-time systems, rather than offline analysis – hence some of the design choices noted herein.

1 INTRODUCTION

Adaptive whitening, described in [7], is a preprocessing technique for STFT-based onset detectors, which can improve the performance of onset detectors by compensating for spectral rolloff and dynamic variability. This MIREX submission makes use of adaptive whitening in conjunction with a range of onset detection functions (ODFs) similar to those described in [7].

This work follows on in particular from the work of Brossier on real-time onset detection [2, 3], but also from [4] and the work summarised in [1].

1.1 Real-time note

The orientation of our research is towards real-time processing, and the present work is developed in that context. Hence the onset detectors (and the adaptive whitening procedure) are all causal and quite efficient. The real-time context, in particular the desire for fast reaction, also motivates the small STFT frame size and the onset selection algorithm used (triggering rather than peak-picking – see below).

Performance improvements may be possible given the non-real-time context of the MIREX contest but these have not been investigated.

2 ADAPTIVE WHITENING

The adaptive whitening algorithm can be expressed as

$$P_{n,k} = \begin{cases} \max(|S_{n,k}|, r, mP_{n-1,k}) & \text{if } n > 0, \\ \max(|S_{n,k}|, r) & \text{otherwise.} \end{cases} \quad (1)$$

$$S_{n,k} \leftarrow \frac{S_{n,k}}{P_{n,k}} \quad (2)$$

for $n \geq 0$, where m is the memory coefficient, r the floor parameter, and $S_{n,k}$ the value of the complex STFT at frame index n and frequency bin index k.

The algorithm is an adaptive process that aims to whiten the signal in the sense of bringing the magnitude of each frequency band into a similar dynamic range; hence the term *adaptive whitening*. As discussed in [7], the dynamic range modifications in the adaptive whitening algorithm are intended to compensate for spectral roll-off, allowing information from high-frequency bands to contribute towards onset detection to a similar extent as that from low-frequency bands; but also to mitigate the effects of strongly-varying dynamics often found in music signals. This compensation is performed adaptively, rather than in fixed fashion as embodied in (e.g.) the high-frequency content measure or pre-emphasis filtering.

3 ANALYSIS PROCEDURE

The analysis is outlined in figure 1. Audio files are analysed using STFT frames of size 512, with 50% overlap between frames and Welch windowing. The adaptive whitening procedure is applied to each STFT frame before the frame is passed to the main onset detection algorithm.

To select onsets, many onset detectors apply a peak-picking procedure after the ODF has been derived, but this work applies a simpler triggering process. An onset is said to be detected when the ODF value rises above the median of recent ODF values (here, the median is over the most recent 11 frames).

4 ONSET DETECTION FUNCTIONS

The following ODFs are included in the submission:

1. Power (Pow)
2. Sum of magnitudes (MS)

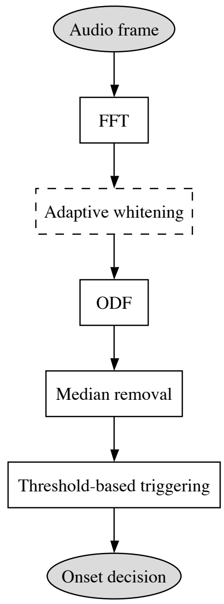

Figure 1. Block diagram of the onset detector used. "ODF" can be any of the ODFs used.

3. Complex deviation (CD)
4. Rectified complex deviation (RCD)
5. Phase deviation (PD)
6. Weighted phase deviation (WPD)
7. Modified Kullback-Leibler divergence (MKL)

The power ODF simply uses the instantaneous power measurement (sum of squared magnitudes). For definitions of ODFs 3–6 see [4]. The Modified Kullback-Leibler divergence algorithm is as presented by Brossier [2, section 2.3]:

$$\mathrm{MKL}_n = \sum_{k=0}^{K} \log\left(1 + \frac{|S_{n,k}|}{|S_{n-1,k}| + \epsilon}\right) \tag{3}$$

where $S_{n,k}$ represents the STFT bin at frequency k for frame n. ϵ is added to the calculation to avoid large variations when very low energy levels are encountered.

For the ϵ coefficient we have tested a number of values; we found $\epsilon = 0.01$ to work much better than the lower values such as 10^{-6} recommended by Brossier [2]. (As with r above, the scaling of ϵ is related to the overall signal amplitude.)

5 IMPLEMENTATION

In [7] the onset detectors were implemented as plugins for SuperCollider [6]. We have "libified" the functions from those plugins, converting them into a self-contained onset-detection library written in C. A simple C executable using this library (plus libsndfile [1] and FFTW [5]) is used in the present MIREX submission.

6 MIREX DATA

For the sake of fairness it must be pointed out that the first author has been working with the dataset used for the MIREX Onset Detection contest; and, while the present work is not directly optimised or trained towards that dataset, the onset detector has been applied to subsections of the dataset during development.

References

[1] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, 2005.

[2] P. M. Brossier. *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis, Queen Mary, University of London, August 2006.

[3] P. M. Brossier, J. P. Bello, and M. D. Plumbley. Real-time temporal segmentation of note objects in music signals. In *Proc. International Computer Music Conference (ICMC'04)*, pages 458–461, 2004.

[4] S. Dixon. Onset detection revisited. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 133–137, Montreal, Quebec, Canada, Sept. 18–20, 2006.

[5] M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005.

[6] J. McCartney. Rethinking the computer music language: SuperCollider. *Computer Music Journal*, 26(4):61–68, 2002.

[7] D. Stowell and M. D. Plumbley. Adaptive whitening for improved real-time audio onset detection. In *Proceedings of the International Computer Music Conference (ICMC'07)*, August 2007.

[1] http://www.mega-nerd.com/libsndfile/