INFLUENCE OF PEAK SELECTION METHODS ON ONSET DETECTION

Carlos Rosão
ISCTE-IUL
L2F/INESC-ID Lisboa
`rosao@l2f.inesc-id.pt`

Ricardo Ribeiro
ISCTE-IUL
L2F/INESC-ID Lisboa
`rdmr@l2f.inesc-id.pt`

David Martins de Matos
IST/UTL
L2F/INESC-ID Lisboa
`david@l2f.inesc-id.pt`

ABSTRACT

Finding the starting time of musical notes in an audio signal, that is, to perform onset detection, is an important task as this information can be used as the basis for high-level musical processing tasks. Many different methods exist to perform onset detection. However their results depend on a Peak Selection step that makes the decision whether an onset is present at some point in time. In this paper we review a number of different Peak Selection methods and compare their influence in the performance of different onset detection methods and on 4 distinct onset classes. Our results show that the post-processing method used deeply influences both positively and negatively the results obtained.

1. INTRODUCTION

In general, music is composed by sounds generated simultaneously by several musical instruments of different kinds [7]. Thus, one can consider the notes played by these musical instruments as the basic unit or syllable for a musical signal [7]. These notes are what allows us humans to clap our hands when listening to a music or whistle/hum the melody of a familiar song [5].

There has been intense research in this area for quite some time, mostly because the information about the starting moments of musical notes can be used as a first step for high-level music processing techniques, such as Chord Estimation, Harmonic Description or Music Genre Classification.

In this paper we are mainly interested in studying how the post-processing part of the onset detection methods, that is, the Peak Selection part in Fig. 1, responsible for deciding whether a point in time is an onset, influences the results obtained. This can be of great help in case one wants to know the more appropriate Onset Detection method – and consequently Peak Selection Method – to use in a particular application.

In the next section, we will present the most common onset detection methods, while in Section 3 we introduce the Peak Selection Methods used. Section 4 describes our
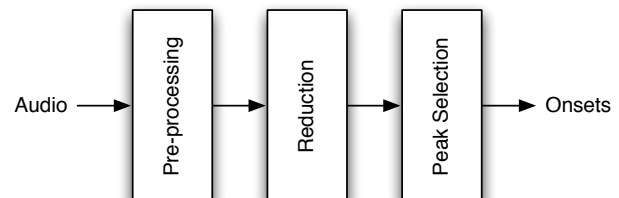
Figure 1. Traditional onset detection work-flow [4].

experiments and discusses the obtained results. The paper ends with final remarks and future work.

2. ONSET DETECTION METHODS

Many Onset Detection Methods have been proposed during the years and most of them follow the general scheme in Fig. 1 which comprises the following steps [1, 4, 5]:

- Pre-processing of the signal in order to highlight its most important properties [1, 4].

- Creation of a Onset Detection Function, also called Onset Strength Signal (OSS) [1], that is, a function whose peaks should correspond to onset times [2].

- Peak Selection, in order to decide which peaks in the Onset Detection Function are onsets.

Next, we briefly review the Onset Detection Functions later used to assess the influence of the Peak Selection part of detecting onsets. For a more general overview of onset Onset Detection Functions, check, for instance, [12] and for a thorough comparison of the performances of the different OSS check, for instance [1] or [13].

In order to detect variations in the properties of the audio signal [2], one can create an OSS by lowering the sample rate of the signal without losing relevant information. This a process called Reduction [1].

All the OSS we will explore are based on Spectral Features of the signal. In order to change from the time-domain to the spectral-domain representation of the audio, we make use of the Short-time Fourier Transform (STFT).

High Frequency Content Making use of the fact that typically, when compared to other audio sources, an onset has relative high energy in higher frequencies [1,

---

[1] In this paper we use the terms Onset Detection Function and OSS interchangeably.

11], it is possible to create a Onset Detection Function that weights each STFT bin proportionally to its frequency. This function is called High Frequency Content (HFC).

Spectral Difference Another possibility to define an OSS is to create a function that measures the variation of magnitude between frequency bins [2, 4]. This type of OSS is called Spectral Difference or Spectral Flux (SF).

Phase Deviation One can also look for onsets by searching for irregularities in the phase of consecutive frequency bins [2], and that is what does the Phase Deviation (PD) Onset Detection Function.

It is possible to improve this function by weighting – Weighted Phase Deviation (WPD) – and normalization [2].

Complex Domain It is possible to combine information from the both the energy and phase of the spectrum to create a Complex Domain (CD) function [3]. This kind of function looks for irregularities in the steady-state of the signal [2] .

A possible improvement for this method is to rectify the function so that it ignores offsets and focuses on onsets [2] – Rectified Complex Domain (RCD).

3. PEAK SELECTION METHODS

A function created with any of the methods introduced in Section 2 will typically show well-localized maxima in positions corresponding to onset times [1]. To extract the onset times from the OSS, Peak Selection methods are used that typically include the steps: Post-processing, Thresholding and Peak-picking.

3.1 Post-processing

Post-processing aims at making the Onset Detection Function uniform so that the processes of thresholding and peak-picking will be easier. This process of increasing the uniformity of the Onset Detection Function typically makes use of normalization methods and filters.

The normalization typically works in one of two ways [2, 5]: (i) Subtract the average value of the function from each value, so that the average will be zero and then divide by the maximum value so that the function will be in the interval [-1,1]; (ii) Subtract the average value of the function from each value and then divide by the maximum absolute deviation, so that the average will be 0 and the standard deviation 1.

The filters used are typically low-pass filters [1, 2, 5], which, in general, select low frequencies up to the cut-off frequency (f_c) and attenuate frequencies higher than f_c [14] and can be defined as

$$y_i = \alpha x_i + (1 - \alpha)y_{i-1} \qquad (1)$$

where α is the smoothing factor.

3.2 Thresholding

In order to separate event-related from non-event-related peaks in the post-processed Onset Detection Function, d, it is common to build a threshold [1].

One can define a constant threshold [8], δ, although this type of threshold is not appropriate, because it does not consider the great dynamics common in a musical signal, leading to weak results [1]. It is much more common to use adaptive thresholds [1, 2, 5]. An adaptive threshold can be constructed in several ways. The best way to overcome problems when facing music pieces with great dynamic change is to build a threshold function based on the local mean (Eq. 2) or local median (Eq. 3) of the Onset Detection Function, d [6].

$$\tilde{\delta}(n) = \delta + \lambda \operatorname{mean}(|d(n - M)|, ..., |d(n + M)|) \quad (2)$$

$$\tilde{\delta}(n) = \delta + \lambda \operatorname{median}(|d(n - M)|, ..., |d(n + M)|) \quad (3)$$

Where λ and δ are positive constants, that can be tweaked, and M is the size of a window around each of the points of the Onset Detection Function.

3.3 Peak-picking

After building a threshold function, one must choose which values of the Onset Detection Function that are larger than the threshold correspond to onsets.

One can consider every value greater than the threshold ($w = 0$ in the following equation) as an onset, or one can add the condition that it must be a local maximum ($w > 0$) [2, 4] (where w is a tweakable parameter that corresponds to the size of a window around the value):

$$o(n) = \begin{cases} 1 & \text{if } d(n) > \tilde{\delta}(n) \\ & \text{and } \quad d(n - w) \leq d(n) \leq d(n + w), \\ \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

4. RESULTS

In this section we will present the evaluation methods and dataset used as well as discuss the results obtained.

4.1 Evaluation Methods

When evaluating onset detection methods, the most common criterion is the F-measure, that is defined in Eq. 5.

$$\text{F-measure} = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2\,P\,R}{P + R} \qquad (5)$$

With Precision, P, and Recall, R, which can be computed in terms of the False Positive (FP), True Positive (TP) and False Negative (FN). In the particular case of onset detection, one can interpret the TP as the correctly detected onsets, the FP as falsely detected onsets and the FN as onsets that were not detected.

The Precision, that is, the fraction of retrieved instances that are relevant is defined in Eq. 6.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (6)$$

On the other hand, the Recall, that is, the fraction of relevant instances that are retrieved, is obtained by Eq. 7.

$$Recall = \frac{TP}{TP + FN} \qquad (7)$$

The Mirex Onset Detection Task specifications [9], and most of the papers in this area, consider onsets detected as TP if they are in a window of $50ms$ around the annotated onset. On the other hand, if more than one detection falls inside the same tolerance window, only one is counted as TP, the others are considered as FP. When a detection is inside the tolerance window of two onset annotations, one TP and one FN are counted. We will evaluate our results according to these specifications.

4.2 Dataset

To run our experiments, we used a dataset built by Bello et al. for [1], referred to as the Bello Dataset.

The Bello Dataset is a hand-labelled and annotated dataset first proposed in [1] and used in several papers, such as [2, 5]. It contains commercial and non-commercial recordings, covering a variety of musical styles and instrumentations, totalling 23 songs and 1065 onsets [1]. The songs are available in WAV format (sample rate 22.050 kHz, mono, 16 bit) and their onset positions (in seconds) in text format.

The recordings of the dataset can be divided in 4 classes, according to the characteristics of their onsets: Complex Mixture (Mix), Pitched Non-Percussive (PNP), Pitched Percussive (PP), and Non-Pitched Percussive (NPP) as shown in Table 1.

	No. Songs	No. Onsets
Mix	7	271
PNP	1	93
PP	9	489
NPP	6	212
Total	23	1065

Table 1. Bello Dataset Structure

One can think of Mix onsets as onsets produced by any polyphonic music where several instruments are playing together, something that happens, for instance, in a rock or pop song. The NPP onsets are the ones typically produced by percussion instruments such as drums or cymbals, while the PP onsets are those that have a percussive characteristic but, nonetheless, still maintain a well defined pitch; this type of onsets appears, for instance, when a piano is playing. Finally, the PNP onsets are those that do not have percussive characteristics and have a very well defined pitch; this category contains onsets from instruments such as bowed strings or wind instruments.

4.3 Experiments

In order to assess the influence of Peak Selection Methods on the results of onset detection, different simulations were run each with a particular Peak Selection Method. These methods were selected because they have been used in recent work [1, 2, 5].

We used the following abbreviations to name the used Peak Selection Methods:

norm Normalize the Onset Detection Function by dividing by the absolute maximum and subtracting the average value, so that the average will be zero.

stdev Normalize the Onset Detection Function by dividing by the maximum standard deviation and subtracting the average value, so that the average will be zero.

mean Create a running mean threshold (Eq. 2).

median Create a running median threshold (Eq. 3).

filter Before normalization, smooth the Onset Detection Function by applying a simple low-pass filter (Eq. 1).

no-filter Do not apply the low-pass filter, that is, do not use smoothing.

local-max Consider as onsets every value in the Onset Detection Function that is larger than zero, larger than the threshold and is a local maximum in a window of 3 samples around it. I.e., use $w = 3$ in Eq. 4.

no-local-max Consider as onset every value greater than the threshold. In other words, use $w = 0$ in Eq. 4.

	A	B	C	D	E
norm	×	×		×	×
stdev			×		
mean		×			
median	×		×	×	×
filter				×	
local-max	×	×	×	×	

Table 4. Components of the Peak Selection Methods A, B, C, D and E.

First we run our experiments with the Peak Selection Method median-norm-no-filter-local-max (A), then we replaced the running mean threshold with a running average threshold with parameter $M = 10$ by running the experiments with the Peak Selection Method mean-norm-no-filter-local-max (B). After that, in order to assess the influence of the type of normalization, we ran the experiments by replacing the norm type of normalization with the stdev type of normalization, that is, using the Peak Selection Method median-stdev-no-filter-local-max (C).

We also tested the influence of a smoothing step before the Peak Selection – with the use of a simple low-pass filter – by running the experiments with the median-norm-filter-local-max (D) Peak Selection Method.

Finally, to test the peak picking algorithm's influence, we ran the experiments without the local maximum condition, that is we used the median-norm-no-filter-no-local-max (E) Peak Selection Method.

	A			B			C			D			E		
OSS	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R
HFC	0.921	0.922	0.920	0.922	0.957	0.901	0.921	0.922	0.920	0.823	0.913	0.766	0.622	0.525	0.798
SF	0.931	0.946	0.926	0.943	0.957	0.937	0.934	0.946	0.932	**0.939**	**0.953**	**0.933**	**0.782**	**0.709**	**0.903**
PD	0.652	0.573	0.819	0.650	0.571	0.819	0.652	0.573	0.819	0.628	0.586	0.749	0.520	0.417	0.893
WPD	0.916	0.959	0.882	0.922	0.933	0.918	0.914	0.945	0.891	0.828	0.900	0.778	0.603	0.507	0.816
CD	**0.947**	**0.978**	**0.923**	**0.946**	**0.987**	**0.913**	**0.943**	**0.970**	**0.923**	0.872	0.931	0.835	0.583	0.482	0.820
RCD	0.933	0.977	0.903	0.933	0.966	0.913	0.936	0.977	0.908	0.909	0.919	0.904	0.419	0.298	0.824

Table 2. Results with P, Precision, F, F-measure and R, Recall, for NPP onsets in the Bello Dataset using all the 5 Peak Selection methods (A, B, C, D, E).

	A			B			C			D			E		
OSS	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R
HFC	0.838	0.846	0.830	0.848	0.829	0.867	0.842	0.846	0.839	0.576	0.607	0.547	0.523	0.437	0.651
SF	**0.961**	**0.968**	**0.954**	**0.965**	**0.978**	**0.953**	**0.961**	**0.969**	**0.954**	**0.876**	**0.878**	**0.874**	**0.893**	**0.868**	**0.921**
PD	0.497	0.410	0.734	0.488	0.414	0.740	0.388	0.278	0.823	0.529	0.323	0.823	0.368	0.256	0.732
WPD	0.810	0.796	0.826	0.811	0.793	0.830	0.811	0.793	0.830	0.470	0.545	0.414	0.666	0.641	0.692
CD	0.883	0.892	0.874	0.899	0.876	0.923	0.903	0.883	0.923	0.441	0.547	0.370	0.543	0.488	0.611
RCD	0.882	0.880	0.883	0.891	0.863	0.920	0.881	0.823	0.947	0.599	0.574	0.625	0.734	0.664	0.820

Table 3. Results with P, Precision, F, F-measure and R, Recall, for PP onsets in the Bello Dataset using all the 5 Peak Selection methods (A, B, C, D, E).

4.4 Discussion

While running the experiments, we fixed the window size of each STFT at 1024 samples (that is 46.4 ms in these 22.05 kHz sampled signals) with a hop size of 50%. The parameters δ and λ were tweaked, in order to obtain the values that maximize the f-measure.

The results obtained by running our experiments with all the Peak Selection Methods described in the previous section are shown in Tables 2, 3, 5 and 6.

In order to compare the methods, we consider as base the results with the Peak Selection Method A and compare all others with this one. First, we will analyse the influence of the Peak Selection Methods on the results obtained for the different onset classes, next, we will analyse the influence of the Peak Selection Methods on each OSS, and, finally, we will make a global balance about the significance of the compared results of the different Peak Selection Methods.

4.4.1 Onset Classes

The differences between running the experiments by using a running-median threshold – Peak Selection Method A – or a running-mean threshold – Peak Selection Method B – have mixed behaviours according to the onset classes. In the NPP and PP classes, the mean gives slightly better results (1pp[2] better) than the median, while it improves for certain OSS it gives worse results for others, but just 1-2pp differences for better or for worse. On the other hand, the running-mean threshold is prone to give worse results by around 2-3pp in the Mix onset class.

To use a normalization based on the maximum standard deviation – Peak Selection Method C – when comparing to a normalization based on the maximum absolute value – Peak Selection Method A – gives mixed behaviours according to the onset classes. In the NPP and PNP onset classes, the results remain almost the same (the changes are less than 1pp) while for the PP the relevant changes

---

[2] pp – percentage point.

are a decrease of around 10pp for the PD function and a performance increase of about 3pp for the HFC and CD functions. When it comes to the Mix onset class, the results for the HFC and PD functions remain just the same, but the other OSS functions have worse f-measure (2-3pp).

When smoothing the Onset Detection Function – Peak Selection Method D – the results become quite different. For the NPP onset class, the SF becomes slightly better (less than 1pp), while for all the other OSS, the results become poorer from 3 to 10pp. In the case of PP onsets, the filter improves about 3pp on the PD function, although it decreases the results significantly (10 to 40pp) for all other OSS. In the PNP onset classes, the behaviour is mixed according to the onset class. We have a positive boost of around 20pp for the PD OSS while for all the other functions the results get worse from 4pp to 30pp. For the Mix onset class, the results get considerably worse for all the OSS.

Finally, when dropping the local maximum condition in the peak picking algorithm – Peak Selection Method E – the results become quite different, but there is a general trend easy to spot: the results get worse for every OSS without exception. In the NPP the results are 15 to 50pp worse, while for the PP the results are 13 to 25pp worse. For PNP onsets, in general, the results are around 30pp worse while for Mix onsets the results vary from 10pp to 30pp worse.

4.4.2 OSS

Moving from running-median threshold to running-mean threshold – Peak Selection Method B – gives, in general, slight improvements for the HFC OSS in all the onset classes, while for the SF OSS the behaviour is mixed. It improves slightly the SF in PP, NPP and PNP onset classes, while decreasing the performance in the Mix class, although these improvements and decreases are very small (1-3pp). We have similar behaviour for the WPD, CD and RCD Onset Detection Functions, with the increases and decreases not going beyond 3pp. In the case of the PD OSS, the re-

OSS	A			B			C			D			E		
	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R
HFC	0.553	0.519	0.591	0.552	0.519	0.591	0.553	0.519	0.591	0.405	0.471	0.355	0.358	0.242	0.688
SF	**0.911**	**0.888**	**0.935**	**0.915**	**0.858**	**0.978**	**0.914**	**0.914**	**0.914**	**0.869**	**0.847**	**0.892**	**0.696**	**0.595**	**0.839**
PD	0.615	0.479	0.860	0.615	0.479	0.860	0.615	0.479	0.860	0.803	0.770	0.839	0.184	0.101	1
WPD	0.660	0.602	0.731	0.670	0.626	0.720	0.670	0.626	0.720	0.465	0.506	0.430	0.463	0.343	0.710
CD	0.684	0.650	0.720	0.680	0.644	0.720	0.677	0.657	0.699	0.388	0.444	0.344	0.409	0.295	0.667
RCD	0.808	0.745	0.882	0.808	0.745	0.882	0.808	0.745	0.882	0.503	0.500	0.505	0.562	0.425	0.828

Table 5. Results with P, Precision, F, F-measure and R, Recall, for PNP onsets in the Bello Dataset using all the 5 Peak Selection methods (A, B, C, D, E).

OSS	A			B			C			D			E		
	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R
HFC	0.812	0.753	0.881	0.814	0.757	0.881	0.814	0.757	0.881	0.597	0.686	0.528	0.512	0.435	0.626
SF	**0.880**	**0.922**	**0.842**	**0.867**	**0.895**	**0.841**	**0.867**	**0.889**	**0.846**	**0.853**	**0.844**	**0.863**	**0.679**	**0.693**	**0.665**
PD	0.540	0.396	0.851	0.540	0.403	0.818	0.544	0.409	0.808	0.491	0.373	0.718	0.458	0.337	0.713
WPD	0.832	0.762	0.811	0.801	0.797	0.806	0.809	0.791	0.822	0.587	0.630	0.564	0.557	0.625	0.505
CD	0.866	0.798	0.854	0.844	0.807	0.870	0.843	0.792	0.881	0.541	0.586	0.518	0.522	0.545	0.509
RCD	0.824	0.823	0.770	0.795	0.818	0.761	0.814	0.814	0.803	0.715	0.680	0.745	0.650	0.652	0.653

Table 6. Results with P, Precision, F, F-measure and R, Recall, for Mix onsets in the Bello Dataset using all the 5 Peak Selection methods (A, B, C, D, E).

sults are quite similar for all the onset classes.

By using a normalization based on the maximum standard deviation – Peak Selection Method C – the results are not very different from the results obtained by using a normalization based on the maximum absolute value – Peak Selection Method A. In the case of the HFC, SF, and RCD, we obtain practically the same results (they change by no more than 1pp) for all the onset classes. In the case of the PD OSS, we have losses of about 10pp for the PP onset class but for the other classes the results remain basically the same (they change by less than 1pp). For the WPD and CD functions the behaviour is mixed, that is, for some onset classes the results improve while for others the results get poorer, although the magnitude of the changes in this OSS is less than 2pp, which means that the changes are not very significant. This Peak Selection Method improves the CD in the PP class, but makes its results worse in the PNP and Mix classes. On the other hand, it improves the WPD in the PNP class, but makes it worse in the Mix class.

The use of a smoothing filter on the Onset Detection Function – Peak Selection Method D – causes the results, in general, to be much different than the results obtained with the Peak Selection Method A. For the HFC OSS, the results decrease from 10 to 25pp and for the SF the tendency is the same, except that for the NPP onset class the results improve slightly (less than 1pp) and the global losses are not so pronounced: they reach at most 9pp. In the case of the PD function we obtain mixed behaviour: for the NPP and Mix onsets the results are 2.5 and 5pp worse respectively while for the PP onsets the results improve by 3pp and for the PNP we have a 20pp improvement. The results get about 2 to 34pp and 7.5 to 44pp worse for the WPD and CD OSS respectively, while for the RCD OSS the results remain similar for NPP class, but get 9 to 30pp worse for the other onset classes. The filter has some kind of "good" effect only on the PD OSS, maybe because this kind of function is the most irregular and the filter brings some positive uniformity, and on the other OSS one obtains an excess of uniformity with the filter, decreasing the

precision of the OSS.

Dropping the local maximum condition in the peak picking algorithm – Peak Selection Method E – makes, in general, the results be much worse than the results of the Peak Selection Method A. For the HFC the results are all around 30pp worse while the results can be to 20pp worse for the SF, 40pp worse for the PD and to 34pp worse for the WPD. For the complex domain family, the results can be to 40pp worse for the CD and 50pp worse for the RCD.

4.4.3 Balance

Having in mind the discussion of the two previous subsections, we can make a global balance. First of all, in general, the differences between the results obtained by applying a running mean and a running median threshold are not statistically significant ($W = 291, p = 0.959$ in the Wilcoxon signed rank sum test with continuity correction [3]) and they are dependent upon the particular onset class and OSS, which implies that for certain applications that need just a certain type of onsets, one specific type of threshold can be chosen in favour of the other.

Concerning the normalization methods, the differences between the results obtained with the two kinds of normalization used are not statistically significant ($W = 290, p = 0.975$ in the Wilcoxon signed rank sum test with continuity correction).

On the other hand, the results obtained by the usage of a smoothing filter get significantly poorer ($W = 427, p = 0.004$ in the Wilcoxon signed rank sum test with continuity correction) in most of the cases, except for the single case of the PD OSS. This means that one should not use a smoothing filter at all (except maybe for the single case of the PD function) or try to test a different filter from the one used in this study.

Finally, not using the local maximum condition makes the results get significantly poorer ($W = 500, p < 0.001$ in the Wilcoxon signed rank sum test with continuity cor-

---

[3] All statistical tests were obtained using R [10].

rection), which means that one should really use the local maximum condition.

5. CONCLUSIONS

In this paper we have compared the influence of 5 distinct Peak Selection Methods on the performance of some of the most common onset detection methods. Our comparison focused on both the influence of the peak selection on each particular OSS but also on the influence of the results in each onset classes.

We have found that, in general, the Peak Selection Method used can be of great influence on the results obtained, but not all of them have the same magnitude of influence. Globally, the influence of using a running-mean or running-average threshold and of using a normalization based on the maximum absolute value or on the maximum standard deviation is quite small (at best around 3-4pp) and can be both positive or negative, depending on the cases. On the other hand using a low-pass filter as a first smoothing step and not using a local maximum condition as final step can be of great negative influence, sometimes worse by 50pp.

We also noticed that, globally, the SF OSS is the most robust to Peak Selection changes, and the PD is the most susceptible to changes.

In the future this work can be extended by adding a few Onset Detection methods to the comparison and also by testing more Peak Selection Methods. One possibility is to add more types of filters to the smoothing to see if the negative influence continues or is just something related to the filter we used. We also intend to check if these conclusions apply to a larger dataset.

6. ACKNOWLEDGEMENTS

7. 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] S. Dixon. Onset Detection Revisited. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 133–137, September 2006.

[3] C. Duxbury, J.P. Bello, M. Davies, and M. Sandler. A combined phase and amplitude based approach to onset detection for audio segmentation. In *Proc. 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-03)*, pages 275–280, Singapore, 2003. World Scientific Publishing Co. Pte. Ltd.

[4] F. Eyben, S. Böck, B. Schuller, and A. Graves. Universal Onset Detection with Bidirectional Long Short-Term Memory Neural Networks. In *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 589–594, 2010.

[5] A Holzapfel, Y Stylianou, A C Gedik, and B Bozkurt. Three Dimensions of Pitched Instrument Onset Detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1517–1527, August 2010.

[6] I. Kauppinen. Methods for detecting impulsive noise in speech and audio signals. In *14th International Conf. on Digital Signal Processing Proc. DSP 2002 (Cat. No.02TH8628)*, volume 2, pages 967–970. IEEE.

[7] A. Klapuri and M. Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, 2006.

[8] A.P. Klapuri, A.J. Eronen, and J.T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):342–355, 2006.

[9] MIREX. Mirex 2011: Audio onset detection task. http://www.music-ir.org/mirex/wiki/2011:Audio_Onset_Detection, May 2011.

[10] R Development Core Team. R: A language and environment for statistical computing. 2008. ISBN 3-900051-07-0.

[11] X. Rodet and F. Jaillet. Detection and modeling of fast attack transients. In *Proc. of the International Computer Music Conference*, pages 30–33, 2001.

[12] C. Rosão and R. Ribeiro. Trends in Onset Detection. In *Proc. of the 2011 Workshop on Open Source and Design of Communication*, pages 75–81. ACM, 2011.

[13] C. Rosão, R. Ribeiro, and D. Martins de Matos. Comparing Onset Detection Methods Based on Spectral Features. In *Proc. of the 2012 Workshop on Open Source and Design of Communication*. ACM, 2012.

[14] U. Zölzer, X. Amatriain, D. Arfib, J. Bonada, G. De Poli, P. Dutilleux, G. Evangelista, F. Keiler, A. Loscos, D. Rocchesso, M. Sandler, X. Serra, and T. Todoroff. *DAFX:Digital Audio Effects*. Wiley, 2002.