LOCAL GROUP DELAY BASED VIBRATO AND TREMOLO SUPPRESSION FOR ONSET DETECTION

Sebastian Böck and Gerhard Widmer
Department of Computational Perception
Johannes Kepler University, Linz, Austria
sebastian.boeck@jku.at

ABSTRACT

In this paper we present a new vibrato and tremolo suppression technique for onset detection. It weights the differences of the magnitude spectrogram used for the calculation of the spectral flux onset detection function on the basis of the local group delay information. With this weighting technique applied, the onset detection function is able to reliably distinguish between genuine onsets and spectral energy peaks originating from vibrato or tremolo present in the signal and lowers the number of false positive detections considerably. Especially in cases of music with numerous vibratos and tremolos (e.g. opera singing or string performances) the number of false positive detections can be reduced by up to 50% without missing any additional events. Performance is evaluated and compared to current state-of-the-art algorithms using three different datasets comprising mixed audio material (25,927 onsets), violin recordings (7,677 onsets) and solo voice recordings of operas (1,448 onsets).

1. INTRODUCTION AND RELATED WORK

Onset detection is the process of finding the starting points of all musically relevant events in an audio performance. While the detection of percussive onsets can be considered a solved problem, [1] softer onsets, vibrato and tremolo are still a major challenge for existing algorithms.

Soft onsets (e.g. bowed string or woodwind instruments) have a long attack phase with a slow rise in energy, thus energy or magnitude-based approaches are not the best fit to detect these sort of onsets. In the past, special algorithms have been proposed to solve the problem of soft onsets by incorporating (additionally) phase [3, 4, 10] or pitch information [9, 14, 15] or a combination thereof [12] to overcome the shortcomings of energy or magnitude-based onset detection algorithms. However, advances in magnitude-based methods [6] show that these methods are now on par with the before-mentioned methods but outperform them on all sorts of percussive audio material.

The current state-of-the-art methods for online [5] and offline [11] onset detection are based on a probabilistic model and incorporate a recurrent neural network with the spectral magnitude and its first time derivative as input features. Especially the offline variant *OnsetDetector* shows superior performance on all sorts of signals [1]. Because of its bidirectional architecture, it is able to model the context of an onset to both detect barely dicernible onsets in complex mixes and suppress events which are erroneously considered as onsets by other algorithms.

Vibrato is an artistic effect commonly used in classical music and can be sung or played by (mostly) string instruments. It reflects a periodic change of the played or sung frequency of the note. Vibrato is technically characterized by the amount of pitch variation (e.g. $\pm$ a semitone for string instruments and up to a complete tone in operas) and the frequency with which the pitch changes over time (e.g. 6 Hz). It is sometimes used synonymously as a combination with another effect: the tremolo, which describes the changes in volume of the note. Because it is technically hard for a human musician to play pure vibratos or tremolos, usually both effects are performed simultaneously. The resulting fluctuations in loudness and frequency make it very difficult for onset detection algorithms to distinguish correctly between new note onsets and an intended variation of the note.

So far only a few publications have addressed the problem of spuriously detected onsets music containing vibrato and tremolo. Collins [9] uses a vibrato suppression stage in his pitch-based onset detection method, which first identifies vibrato regions that fluctuate at most one semitone around the center frequency and collects the extrema in a list. The region is expanded gradually in time to cover the whole duration of the vibrato. After having identified the complete extent of the vibrato, all values within this window are replaced by the mean of the extrema list. The onset detection function is based on the concept of stable pitches and uses the change in pitches as cues for new onsets.

Schleusing et al. [14] deploy a system based on the inverse correlation of N consecutive spectral frames centered around the current location. Regions of stable pitch lead to low inverse correlation values, and pitch changes result in peaks in the detection function. To suppress vibrato they deploy a warp compensation which cancels out

[1] F-measure values > 0.95 as obtained with state-of-the-art onset detection algorithms [1] can be considered to have solved the problem.

small pitch changes within the considered window, leaving genuine onsets mostly untouched.

Recent research [7] applies a maximum filter to suppress vibrato in audio signals. This method operates in the spectral domain; specifically it only considers the magnitude spectrogram without incorporating any phase information. Like the common spectral flux algorithm [13] it relies on the detection of positive changes in the energy over time, but instead of calculating the difference between the same frequency bin for the current and previous frames, it includes a special magnitude trajectory tracking stage which is able to suppress spurious positive energy fragments.

Still, all algorithms (apart from those relying solely on phase information) suffer from loudness variations, which mostly originate from the tremolo effect. This paper addresses this problem by incorporating the phase – more specifically the local group delay (LGD) information – to determine steady tones and suppress the spurious loudness variations accordingly.

2. PROPOSED METHOD

Incorporating phase information is only feasible if each frequency bin of the spectrogram is considered separately as in the methods described in [3, 4, 10]. However, these methods have proven to perform poorly compared to current state-of-the-art algorithms [6]. Thus, our method is based on the recently proposed *SuperFlux* [7] algorithm, which is an enhanced version on the common spectral flux algorithm [13]. It is already significantly less sensitive to frequency variations caused by vibrato, but adding a special local group delay based weighting technique to the difference calculation step, makes this method even more robust against loudness variations of steady tones, e.g., those caused by tremolo.

2.1 SuperFlux

The system performs a frame-wise processing of the audio signal (sample rate 44,1 kHz). The signal is divided into overlapping chunks of length $N = 2048$ samples and each frame is weighted with a Hann window of the same length before being transformed to the spectral domain via the discrete Fourier transform (DFT). Two adjacent frames are located 220.5 samples apart, resulting in a resolution of 200 frames per second, which allows reportin of onsets to within 5 ms.

It has been found advantageous [6] to first filter the resulting magnitude spectrogram $|X(n,k)|$ (n denotes the frame number and k the frequency bin index) with a filterbank $F(k,m)$ (with m being the filter band number) before being processed further. The filterbank has $M = 138$ filters aligned equally on the logarithmic frequency scale with quarter-tone spacing. To better match the human perception of loudness, the resulting filtered spectrogram $X_F(n,m)$ is then transferred to a logarithmic magnitude scale, denoted $X_{L,F}(n,m)$ hereafter. Instead of calculating the bin-wise difference to the previous frame of the same logarithmic filtered spectrogram, a maximum filter along the frequency axis is applied (i.e. the value of a bin is set to the maximum of the same bin and its direct neighbors on the frequency axis) and the difference is calculated with respect to the μ-th previous frame of this maximum filtered spectrogram $X^{max}_{L,F}(n,m)$ resulting in the following equation for the difference calculation stage:

$$D(n,m) = X_{L,F}(n,m) - X^{max}_{L,F}(n-\mu,m) \quad (1)$$

The parameter μ depends on the frame rate f_r, which is set to 200 fps, resulting in $\mu = 2$ frames. The *SuperFlux* onset detection function is then defined as the sum of all positive differences:

$$SF(n) = \sum_{m=1}^{m=M} H\left(D(n,m)\right) \quad (2)$$

with $H(x) = \frac{x+|x|}{2}$ being the half-wave rectifier function.

The positive effect of these measures can be seen clearly in Figures 1a to 1c, which depict a 4 second recording of a violin played with vibrato and tremolo. However, there are still some spurious positive energy fragments left, which can be eliminated with the approach described in the next section. For a more detailed description of the *SuperFlux* algorithm, please refer to [7].

2.2 Local Group Delay based difference weighting

Using solely the magnitude information of the spectrogram enables onset detection algorithms to detect most onsets reliably, but also makes them susceptible to all kinds of loudness variations of steady tones. Using the phase as an additional source of information helps to lower the impact of these loudness variations. However, the main problem of incorporating the phase information is that it can only be combined easily with the magnitude spectrogram if all frequency bins of the STFT are considered individually. But since filtering the magnitude spectrogram with a filterbank (i.e. merging several frequency bins into a single one) previous to the difference calculation yields much better performance for almost all kinds of audio signals [6], the phase information of constituent frequency bins of a filter band have to be combined such that phase can be used in conjunction with the filtered spectrogram.

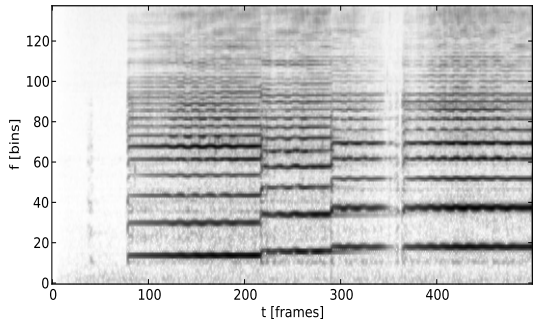
We investigated different approaches for combining the phase information of several frequency bands into one, and propose the following simple but effective solution. Given the phase ϕ of the complex spectrogram X by:

$$\phi(n,k) = \text{angle}\left(X(n,k)\right), \quad (3)$$

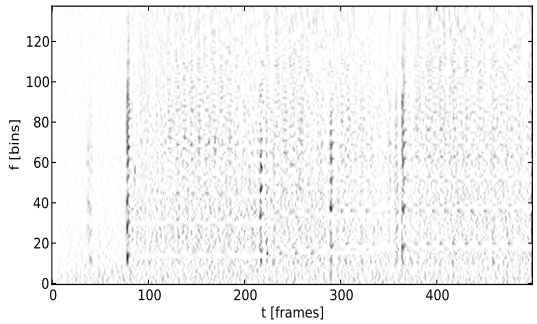
we can estimate the local group delay (LGD) of the spectrogram as:

$$LGD(n,k) = \phi^*(n,k) - \phi^*(n,k-1), \quad (4)$$

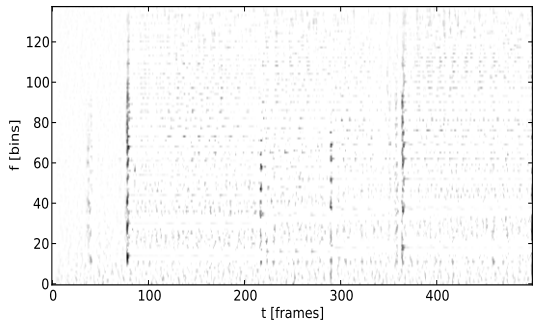
with ϕ^* defined as the 2π-unwrapped (over the frequency axis) phase. The local group delay gives information as where the gravitational centre of the magnitude is located. The spectrogram reassignment method [2] uses this information to gather a sharpened (reassigned) representation



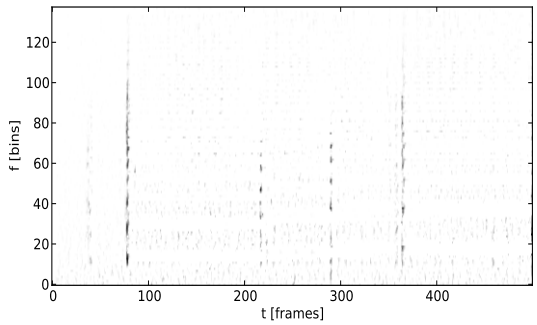
(a) magnitude spectrogram



(b) classical bin-wise positive difference



(c) maximum filtering trajectory tracking based difference



(d) local group delay based difference weighting

Figure 1: *(a) logarithmic magnitude spectrogram of a 5s violin played with vibrato and tremolo, (b) the positive differences calculated as in the spectral flux algorithm, (c) with applied maximum filtering as in [7] and (d) the proposed local group delay based difference weighting approach .*

of the magnitude spectrogram. Although this representation is more exact, the process leads to areas with lower magnitudes. The reassigned spectrogram looks a bit like a "scattered" version of the well known magnitude spectrogram. Thus, using this representation directly to calculate the spectral flux showed worse performance, mostly because of lots of smaller energy peaks, which we are trying to avoid.

Instead of using the local group delay information to relocate the magnitudes of the spectrogram, the information can be interpreted in a different way: regions with values close to zero indicate stable tones (or percussive sounds if they are aligned along the frequency axis) and regions with absolute values greater than zero indicate a possible onset. Holzapfel et al. [12] use the average of all local group delay values along the frequency axis as a feature for their onset detection function. Instead of averaging the individual values, we determine the local minimum within each band of the filterbank $F(k, m)$ for the *SuperFlux* calculation, and use these values as a weighting function.

Care has to be taken that the individual filters of the filterbank do not cover too many frequency bins, as the likelihood that there is a local group delay minimum that does not belong to any steady tone increases accordingly. Filterbanks with 24 filters per octave yielded good results for all kinds of music material. The higher the expected fluctuations in frequency, the lower should be the chosen number of filter bands. However, the fewer filter bands used, the wider the individual filter bands become, and in turn, this impacts the performance on percussive onsets. Percussive onsets have low local group delay values over a broad range of the frequency axis, thus applying the local minimum as a weighting would "erase" almost all percussive onsets.

To lower the impact of local group delay weighting on percussive sounds, we first apply a maximum filter over time which covers the range of 15 ms. For a frame rate of $f_r = 200$ fps, this equals to three frames and results in a temporal maximum filtered version of the LGD spectrogram:

$$LGD^*(n, k) = \max\left(|LGD(n-1:n+1, k)|\right) \quad (5)$$

After this first filtering step, we get the final local group delay based weighting by applying the previously described minimum filter, which sets the value of a bin to the local minimum of the region defined by the filter band:

$$W(n, m) = \min\left(LGD^*(n, k_{L(m)} : k_{U(m)})\right) \quad (6)$$

with $k_{L(m)}$ representing the lower frequency bin index of the filter band m of the filterbank $F(k, m)$, and $k_{U(m)}$ the upper bound respectively. This function is then used to weight the difference of the *SuperFlux* (cf. Equation 1), resulting in the modified detection function:

$$SF^*(n) = \sum_{m=1}^{m=M} H\left(D(n, m)\right) \cdot W(n, m) \quad (7)$$

with $H(x) = \frac{x+|x|}{2}$ being the half-wave rectifier function, n the frame number and m the frequency bin index. The '$\cdot$' operator denotes the element wise multiplication of the two matrices.

The effect of all proposed measures can be seen in Figure 1. Compared to the standard spectral flux implementation (1b), the difference with applied maximum filtering trajectory tracking (1c) already shows fewer positive energy components, which are further reduced by the proposed method, as can be seen in (1d). Figure 2 shows the sums of the positive differences. It is evident that the new approach lowers the overall noise in regions with vibrato and tremolo but keeps very sharp peaks at the onset positions.

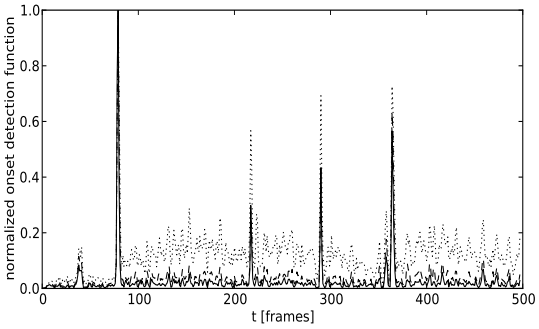


Figure 2: *Spectral flux sum of the differences shown in Figure 1. The simple filtered spectral flux is shown as dotted line, the SuperFlux as dashed line, and the proposed local group delay based difference weighting approach as solid line.*

It should be mentioned that the same weighting technique could be used for unfiltered magnitude spectrograms (i.e. the original spectral flux implementation). Instead of using the local maximum of all frequencies of a filter band, only the same frequency bin and its direct neighbors should be considered. Although the same positive impact on signals containing vibrato and tremolo can be observed, the overall performance compared to the filtered variants of the spectral flux (e.g. the *LogFiltSpecFlux* [6] or the *SuperFlux* [7]) is much lower, especially for polyphonic music.

2.3 Peak-picking

For selecting the final onsets of the weighted *SuperFlux* detection function we use the same peak-picking method as in [7]. Since the new onset detection function $SF^*(n)$ has a lower noise floor and shows sharper peaks than the original implementation (Equation 2), we had to alter the parameters for the peak-picking method used in [7]. A frame n of the onset detection function $SF^*(n)$ is selected as an onset if it fulfills the following three conditions:

1. $SF^*(n) = \max\left(SF^*(n-\omega_1 : n+\omega_2)\right)$
2. $SF^*(n) \geq \mathrm{m}ean(SF^*(n-\omega_3 : n+\omega_4)) + \delta$
3. $n - n_{previous\,onset} > \omega_5$

where δ is the tunable threshold. The other parameters were chosen to yield the best performance on the complete dataset. $\omega_1 = 30\,\mathrm{ms}$, $\omega_2 = 30\,\mathrm{ms}$, $\omega_3 = 100\,\mathrm{ms}$, $\omega_4 = 70\,\mathrm{ms}$ and the combination width parameter $\omega_5 = 30\,\mathrm{ms}$ showed good overall results. Parameter values must be converted to frames depending on the frame-rate f_r used.

3. EVALUATION

For the evaluation of the algorithm, different datasets and settings have been used to allow highest comparability with previous publications.

3.1 Performance measures and evaluation settings

For evaluating the performance of onset detection methods, commonly Precision, Recall, and F-measure are used. If a detected onset is within the evaluation window around an annotated ground truth onset location, it is considered a correctly identified onset. But every detected onset can only match once, thus any detected onset within the evaluation window of two different annotated onsets counts as one true positive and one false negative (a missed onset). The same applies to annotations, i.e. all additionally reported onsets within the evaluation window of an annotation are counted as false positive detections. In order to keep the comparability with other results, we match the evaluation parameters as follows:

Our standard setting is the one used in [6], which combines all annotated onsets within 30 ms to a single onset and uses an evaluation window of $\pm$ 25 ms to identify correctly detected onsets. Thus the combination width parameter ω_5 of our peak-picking method is set to 30 ms as well.

The second set of parameters (denoted with an asterisk in Table 1) uses the same settings as in [14], where all onsets within 50 ms are combined (i.e. $\omega_5 = 50\,\mathrm{ms}$) and an evaluation window of $\pm$ 70 ms is used.

Unless otherwise noted, all given results are obtained by swiping the threshold parameter δ of the peak-picking stage and choosing the value that maximizes the F-measure on the respective dataset.

3.2 Datasets

For comparison with the former state-of-the-art algorithm for pitched non-percussive music, the dataset from [14] is used. Unfortunately not all sound files and annotations could be used for evaluation, since the authors were only able to provide part of this set. Still, we believe that the achieved results are comparable, because the dataset has over three quarters of the size of the original dataset (7,677 instead of 9,717 onsets) and an identical distribution of the different playing styles (50% contain vibrato, some staccato etc.). This will be called the *Wang* dataset.

To show the ability to suppress tremolo and vibrato present in sung opera vocals, a second dataset introduced in [7] and consisting of solo singing rehearsal recordings of a Haydn opera is used. The set covers both male and female singers and has a total length of 10 minutes containing 1,448 onsets. It is called the *Opera* dataset.

The biggest dataset used for evaluation is that described in [6], which consists mostly of mixed audio material covering different types of musical genres performed on various instruments. It includes the sets used in [3], [12], and [11]. The 321 files have a total length of approximately 102 minutes and have 27,774 annotated onsets (25,927 if all onsets within 30 ms are combined). The main purpose of this set is to show how the new local group delay weighting for the *SuperFlux* algorithm impacts the performance on a general purpose dataset. This dataset is named *Böck*. Based on this set, we build a subset that contain violin and cello recordings played with vibrato and tremolo, but also feature accompaniment instruments. These 16 files have 849 onsets.

3.3 Results & Discussion

Because the local group delay weighting technique is designed especially for audio signals containing mostly vibrato and tremolo, the main focus should be put on the results obtained on the *Wang* and *Opera* datasets. But since we expect that it does not harm the overall performance of the underlying *SuperFlux* algorithm too much when used on other musical signals, the results given on the general purpose *Böck* dataset should not be neglected.

3.3.1 Competitors

Besides the former state-of-the-art algorithm for pitched non-percussive music presented in [14] (for comparison on the *Wang* dataset), we chose the winning submissions of last year's MIREX evaluation [1] for comparison. We consider these submissions to be state-of-the-art, since they achieved the highest F-measure ever measured during the MIREX evaluation.

The *OnsetDetector.2012* is an improved version of the method originally proposed in [11], which shows superior performance in offline scenarios, and represents the group of probabilistic onset detection approaches. Since the *OnsetDetector.2012* was trained on the *Böck* dataset, the results given in Table 3 and 4 for this algorithm were obtained with 8-fold cross-validation and parameters selected solely on the training set. Instead of the *LogFiltSpecFlux* [6] algorithm, we chose the recently proposed *SuperFlux* algorithm [7], which shows better performance on all datasets. The *SuperFlux* algorithm does not use any probabilistic information and thus has much lower computational demands, marking the current upper bound of performance of so-called "simple" algorithms.

Because the onset detection functions of the compared methods show very different shapes and characteristics, and the choice of peak-picking methods and parameters highly influence the final results, we use offline peak-picking only. Since all algorithms yield their best performance in offline mode and are less sensitive to variations of parameters, we consider this a valid choice. Nonetheless, all algorithms can be used in online mode with slightly lower performance.

3.3.2 Wang set

Table 1 shows the performance on violin music for the *Wang* dataset. The new local group delay weighted *SuperFlux* method outperforms all other algorithms with respect to false positive detections by at least 25%. Compared side-by-side with the current state-of-the-art onset detection algorithm, the *OnsetDetctor*, the weighted *SuperFlux* is able to achieve the same level of true positive detections, but improves regarding false positive detections by an impressive 56%.

	TP	FP
OnsetDetector.2012 [11] *	96.5%	15.5%
Schleusing et.al. [14] *	91.2%	9.2%
SuperFlux [7] *	94.7%	9.1%
SuperFlux w/ LGD weighting *	97.0%	6.8%

Table 1: *True and false positive rates of different onset detection algorithms on the Wang dataset. Results for Schleusing's algorithms were taken from [14]. Asterisks mark the evaluation method used in [14].*

Since the recordings in the *Wang* dataset are exclusively solo recordings made in a sound absorbing room and contain only very few polyphonic parts, this result can be seen as the maximum possible performance boost that can be obtained with the local group delay weighting method for this type of music.

3.3.3 Opera set

On the *Opera* dataset with male and female opera rehearsal recordings, the new method also shows its strength and is able to dramatically lower the number of false positive detections. Compared with the original *SuperFlux* implementation, the number of false detections go down from 450 to 221 (which is a reduction by 51%), if the new local group delay based weighting technique is applied. The new method even outperforms the current best-performing probabilistic approach (with respect to F-measure), but it should be noted that the neural network based method was not trained on any opera material.

	P	R	F
OnsetDetector.2012 [11]	0.576	0.777	0.662
SuperFlux [7]	0.672	0.635	0.653
SuperFlux w/ LGD weighting	0.806	0.635	0.711

Table 2: *Precision, Recall and F-measure of different onset detection algorithms on the Opera dataset.*

3.3.4 Böck set

In Table 3 results for the full *Böck* dataset are given. With the new difference weighting scheme, slightly lower performance can be observed. This was expected, since the new approach is tuned specifically towards music with vibrato and tremolo but which otherwise contains only very

few percussive sounds (as present in complex audio mixes like pop songs). It could be argued, that the impressive performance gains achievable for this special type of music justify the small performance penalty on this dataset.

	P	R	F
OnsetDetector.2012 [11]	0.892	0.855	0.873
SuperFlux [7]	0.883	0.793	0.836
SuperFlux w/ LGD weighting	0.873	0.778	0.823

Table 3: *Precision, Recall and F-measure of different onset detection algorithms on the Böck dataset.*

More interesting are the results given in Table 4 for the *strings* subset, which includes pieces with string instrumentation that also feature accompaniment instruments – which make vibrato and tremolo suppression harder. As can be seen, the local group delay weighted *SuperFlux* method also performs slightly worse than the original *SuperFlux* implementation. Thus, it must be concluded that the new weighting scheme is mainly suited for signals which feature numerous vibratos and tremolos but do not contain many other instruments.

	P	R	F
OnsetDetector.2012 [11]	0.834	0.820	0.827
SuperFlux [7]	0.836	0.701	0.762
SuperFlux w/ LGD weighting	0.777	0.710	0.742

Table 4: *Precision, Recall and F-measure of different onset detection algorithms on the strings subset of the Böck dataset using the same parameters as used for the results in Table 3.*

4. CONCLUSIONS

In this paper a new method for vibrato and tremolo suppression with local group delay based spectral weighting was presented. The new weighting scheme can be applied to any spectral flux like onset detection method and is able to reduce the number of false positive detections originating from vibrato and tremolo by up to 50% compared to current state-of-the-art implementations.

For future versions of this weighting technique, the Constant-Q transform could be investigated. Using this transform instead of the Short-Time Fourier Transform would make both the use of a filterbank for the magnitude spectrogram and the rather simple combination technique for the phase information of several frequency bins into one obsolete, but retain the beneficial behavior of this approach.

5. ACKNOWLEDGMENTS

This work is supported by the European Union Seventh Framework Programme FP7 / 2007-2013 through the PHENICX project (grant agreement no. 601166).

6. REFERENCES

[1] MIREX 2012 onset detection results. `http://nema.lis.illinois.edu/nema_out/mirex2012/results/aod/`, 2012, accessed 2013-03-27.

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

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

[4] J.P. Bello, C. Duxbury, M. Davies, and M. Sandler. On the use of phase and energy for musical onset detection in the complex domain. *IEEE Signal Processing Letters*, 11(6):553–556, June 2004.

[5] S. Böck, A. Arzt, F. Krebs, and M. Schedl. Online real-time onset detection with recurrent neural networks. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, York, UK, September 2012.

[6] S. Böck, F. Krebs, and M. Schedl. Evaluating the online capabilities of onset detection methods. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, pages 49–54, Porto, Portugal, October 2012.

[7] S. Böck and G. Widmer. Maximum filter vibrato suppression for onset detection. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, September 2013.

[8] N. Collins. A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions. In *Proceedings of the AES Convention 118*, pages 28–31, Barcelona, Spain, May 2005.

[9] N. Collins. Using a pitch detector for onset detection. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, September 2005.

[10] S. Dixon. Onset detection revisited. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, pages 133–137, Montreal, Quebec, Canada, September 2006.

[11] F. Eyben, S. Böck, B. Schuller, and A. Graves. Universal onset detection with bidirectional long short-term memory neural networks. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 589–594, Utrecht, Netherlands, August 2010.

[12] 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.

[13] P. Masri. *Computer Modeling of Sound for Transformation and Synthesis of Musical Signals*. PhD thesis, University of Bristol, UK, December 1996.

[14] O. Schleusing, B. Zhang, and Y. Wang. Onset detection in pitched non-percussive music using warping-compensated correlation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, pages 117–120, April 2008.

[15] R. Zhou, M. Mattavelli, and G. Zoia. Music onset detection based on resonator time frequency image. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1685–1695, November 2008.