Power-Scaled Spectral Flux and Peak-Valley Group-Delay Methods for Robust Musical Onset Detection

Li Su
CITI, Academia Sinica, Taipei, Taiwan
lisu@citi.sinica.edu.tw

Yi-Hsuan Yang
CITI, Academia Sinica, Taipei, Taiwan
yang@citi.sinica.edu.tw

ABSTRACT

A robust onset detection method has to deal with wide dynamic ranges and diverse transient behaviors prevalent in real-world music signals. This paper presents contributions to robust onset detection by proposing two novel onset detection methods. The first one, termed *power-scaled spectral flux* (PSSF), applies power scaling to the spectral flux to better balance the wide dynamic range in the spectrogram. The second method, called *peak-valley group-delay* (PVGD), enhances the robustness to noise terms by detecting peak-valley pairs from the summed group-delay function to capture the attack-decay envelope. The proposed methods are evaluated on a piano dataset and a diverse dataset of 12 different Western and Turkish instruments. To tackle the problem from a fundamental signal processing perspective, in this study we do not consider advanced methods such as late fusion, multi-band processing, and neural networks. Experimental result shows that the proposed methods yield competitive accuracy for the two datasets, improving the F-score for the former dataset from 0.956 to 0.963, and the F-score for the latter dataset from 0.712 to 0.754, comparing to existing methods.

1. INTRODUCTION

An onset detection functions is generally designed to identify new events in an audio signal by probing the differences in the magnitude, phase angle, complex spectrum or other feature representations. For example, *spectral flux* (SF) computes the temporal differences of magnitude spectra, *phase deviation* (PD) computes the second-order temporal differences of phase angles, whereas *complex domain* (CD) considers simultaneously the differences of both magnitude and phase [1]. Comprehensive overview of the commonly used onset detection and post-processing methods can be found in [2, 3, 4, 5]. In general, SF-based approach is the most popular one in the literature.

The negative of phase slope with respect to frequency, also known as the *group-delay function* (GDF) [6, 7, 8], stands for the distance between the center of the analysis window and the position of the attack-decay wavepacket, which represents an onset event. Therefore, an onset can be detected by the zero-crossing of GDF or the peak of negative GDF derivative [9, 10, 11, 12]. GDF has been found a competitive approach using phase information, especially when multi-band processing is applied [9, 10]. However, studies on GDF have been relatively fewer than SF, possibly due to that phase information is hard to be computationally modeled.

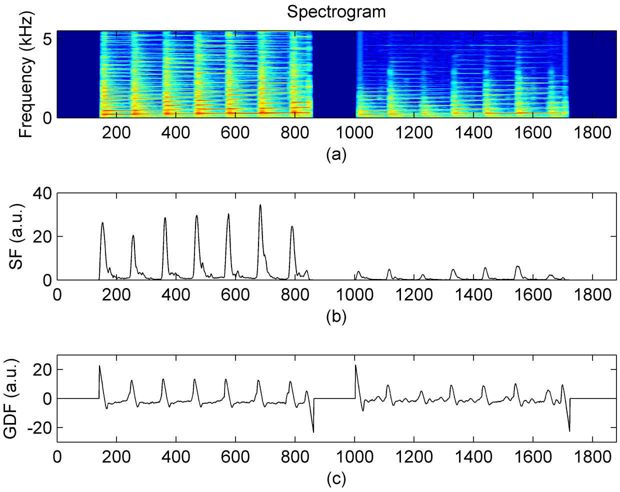

Figure 1. Spectrogram, SF and summed GDF of two piano semitone sequences with different dynamics. Horizontal axis indexes time.

Comparing to GDF, SF is relatively more insensitive to noise, windowing effects and sampling rates, but SF does not perform well for signals with high variation of dynamic range. On the other hand, the GDF is relatively insensitive to changes in signal power, but the performance of phase can be affected by noises and other numerical problems. An example is shown in Fig. 1, which displays the spectrogram, SF and GDF of two succeeding semitone sequences, both from C4 to F4. The two sequences are played in *forte* (f; ‘loud’) and *piano* (p; ‘soft’), respectively. We see that SF is fairly sensitive to dynamics, making it difficult to determine a decision threshold for onset detection. In contrast, GDF is relatively scale-invariant, but it is subject to the noisy terms in the signal and thereby over-emphasizes some unimportant parts. These issues are more pronounced as the type of instruments, the number of playing techniques and the variation in the recording environment that are under consideration increase, which is common in real-world music signals [2, 3, 4, 5].

In this paper, we propose two improved onset detection methods for SF and GDF to circumvent the scaling and robustness issues. Specifically, we add a power-scale parameter to SF to compensate for the musical dynamics, and propose a peak-valley picking method for GDF to deal with transient events (Section 3 and 4). Evaluation on two onset detection datasets [13, 9] validates the effectiveness of the proposed methods over existing methods (Section 5). Although it is interesting to combine the result of SF- and GDF-based methods, we opt for leaving this as a future work as a similar late-fusion approach has been shown effective in [9].

2. BACKGROUND

In this section we introduce some well-known onset detection functions (ODFs) which make use of basic signal properties such as magnitude spectra, complex spectra and phase. We take three baseline ODFs, called spectral flux, weighted phase deviation and the complex domain detection function, which have been found performing will among other basic ODFs [14]. Moreover, difference of group-delay functions is also taken into consideration. Here we use the symbol "Δ" or "$\prime$" to refer to difference or derivative w.r.t time but not frequency (this should not be confused with the definition of GDF, which takes the derivative w.r.t. frequency).

Spectral flux (SF) is arguably the most widely-used ODF. It measures the positive changes in each frequency bin and sums up all these changes within a frame. SF is defined as

$$\mathrm{SF}\left(n\right)=\sum_{k=1}^{N/2}H\left(\left|\mathrm{X}\left(n,k\right)\right|-\left|\mathrm{X}\left(n-1,k\right)\right|\right)\ , \qquad (1)$$

where $H=\left(\mathrm{X}+|\mathrm{X}|\right)/2$ is the half-wave rectifier function. Variants of SF use either l_2-norm formulation [2] or take the logarithm magnitude $\log\left(1+|\mathrm{X}|\right)$ [15, 4].

Weighted phase deviation (WPD) is an improved version of PD [3]. The ODF is obtained by summing up the second-order difference of 2π-unwrapped phase ψ'' for each frame. As WPD is sensitive to noise terms introduced by components with insignificant energy, a magnitude weighting on ψ'' is applied to suppress the insignificant parts [3]:

$$\mathrm{WPD}\left(n\right)=\frac{1}{N}\sum_{k=-N/2}^{N/2-1}\left|\mathrm{X}\left(n,k\right)\psi''\left(n,k\right)\right|. \qquad (2)$$

Complex domain (CD). In complex domain one can jointly incorporate phase and magnitude information instead of processing them separately [1, 3, 2]. CD takes a steady-state "prediction" of the current spectrum $\mathrm{X}_T\left(n,k\right)$, which is evolved from the preceding spectrum and its first-order phase difference ψ':

$$\mathrm{Y}\left(n,k\right)=\left|\mathrm{X}\left(n-1,k\right)\right|e^{j\psi(n-1,k)+j\psi'(n-1,k)}\ , \qquad (3)$$

and then obtains the ODF by calculating the differences in the observed spectrum and the predicted one [3]:

$$\mathrm{CD}\left(n\right)=\sum_{k=-N/2}^{N/2-1}H\left(\left|\mathrm{X}\left(n,k\right)-\mathrm{Y}\left(n,k\right)\right|\right)\ . \qquad (4)$$

Difference of group-delay (ΔGD) represents a simple way to implement an ODF using GDF. It sums up the first-order difference of negative GDF in each frame:

$$\Delta\mathrm{GD}\left(n\right)=-\sum_{k=1}^{N/2}\left(\mathrm{GDF}\left(n,k\right)-\mathrm{GDF}\left(n-1,k\right)\right)\ . \qquad (5)$$

ΔGD can be viewed as a simplification of "ΔGRD" (i.e., difference of auditory group delay) [10], which further processes GDF in several auditory bands separately instead of summing them directly. We will see in the experiments that GDF-based methods are generally better than other phase-based methods such as WPD and CD, possibly due to that the GDF formulation in (10) avoids the requirement of phase unwrapping, which is usually unstable under noisy situations. Details about GDF are introduced in Section 4.1.

3. POWER-SCALED SPECTRAL FLUX (PSSF)

It has been known that a simple log-scale mapping $\log|\mathrm{X}|$ is not applicable as it diverges to negative infinity when $|\mathrm{X}|$ is small. An alternative form $\log\left(1+|\mathrm{X}|\right)$ resolves this issue but weakens the difference of low-energy counterparts at the same time (note that the derivative of $\log\left(1+|\mathrm{X}|\right)$ is strictly bounded between 0 and 1). Therefore, as the dynamic range of the musical signal widens, such a logarithmic form is similar to linear-scaling and shows less advantages.

Power-scale mapping is free from the drawbacks of the two aforementioned logarithmic forms, as a power-scale function $|\mathrm{X}|^p$ for $0<p\leq 1$ has bounded values and unbounded derivatives. The power-scaled variant of SF is defined as

$$\mathrm{PSSF}\left(n\right)=\sum_{k=1}^{N/2}H\left(\left|X\left(n,k\right)\right|^p-\left|X\left(n-1,k\right)\right|^p\right)\ . \qquad (6)$$

The introduction of the power scale p enhances the weaker onsets and suppresses the stronger onsets in a music signal. As a result, PSSF will be more robust for music signals composed of both loud sentences and weak sentences, such as the one shown in Fig. 1(a).

Although p seems like an empirically-determined parameter, there is no need to try all possible values for p because the dynamic ranges of most music signals are limited. Consider an extreme case: the onset strengths of 1 violin and 1,000 violins may differ by about 30 dB; for a common pop music, the dynamic range is usually 6–10 dB, according to some informal studies. Therefore, setting $p=0.5$ should be enough for most cases as this reduces the dynamic range to less than 3 dB (cf. Section 5).

4. PEAK-VALLEY GROUP-DELAY (PVGD)

4.1 Group-delay and onset detection

We begin with an introduction of group-delay and its relation to onset. Consider a musical note as a mixture of

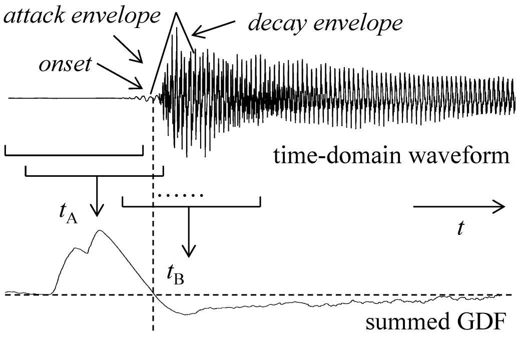

Figure 2. A conceptual diagram on the relation among the time-domain signal, attack-decay wavepacket, analysis window and group-delay.

wavepackets with the corresponding ADSR (attack, decay, sustain, and release) envelopes and their carrier frequencies. It is well-known that GDF describes the delay of a wavepacket, whereas the phase-delay describes the delay of a carrier [16, 8]. Specifically, GDF describes the relative position between the wavepacket and the analysis window. As an onset event is mostly characterized by the attack-decay sub-envelopes, or *transients* [2], it is possible to describe an onset event by GDF.

Fig. 2 shows a band-limited signal modeled with its transient envelope with a series of analysis windows. At time t_A, the window function covers the attack envelope on the right-hand side, and the window function can measure a positive group-delay for this wavepacket at this time. At time t_B the window function covers the end of decay phase in its left-hand side, implying a time advance of analysis window to the wavepacket and therefore a negative group-delay (i.e., beneath the horizontal dashed line). The onset event is thus between the peak and valley of the GDF. For those signals with weak decay envelope like most of the string and wind instruments, the valleys of the GDF also become weak, but still can be identified through peak picking. As the valley becomes weak, the proposed method reduces to ΔGD method (see Section 2). This method is unable to identify the signal with no 'decay' counterpart, such as an ideal Heaviside function, or a *crescendo* note. Note that the onset of a *crescendo* note is also difficult to be detected by other methods.

GDF can be computed directly from the short-time Fourier transform (STFT). Consider a general representation of STFT of a time-domain signal $x(t)$:

$$\begin{aligned} \mathrm{S}_x^h(t,\omega) &= \int_{-\infty}^{\infty} x(\tau)\, h^*(\tau - t)\, e^{-j\omega\tau} d\tau & (7) \\ &= \mathrm{M}_x^h(t,\omega)\, e^{j\Phi_x^h(t,\omega)}\,, & (8) \end{aligned}$$

where $\mathrm{S}_x^h(t,\omega) \in \mathbb{C}$ is the two-dimensional STFT representation on time-frequency plane, $h(t)$ is the window function, $\mathrm{M}_x^h(t,\omega)$ and $\Phi_x^h(t,\omega)$ of Eq. (8) represent the amplitude and phase, respectively. Phase is the imaginary part of the logarithm of Eq. (8):

$$\Phi_x^h(t,\omega) = \mathrm{Im}\left(\log \mathrm{S}_x^h(t,\omega)\right) . \quad (9)$$

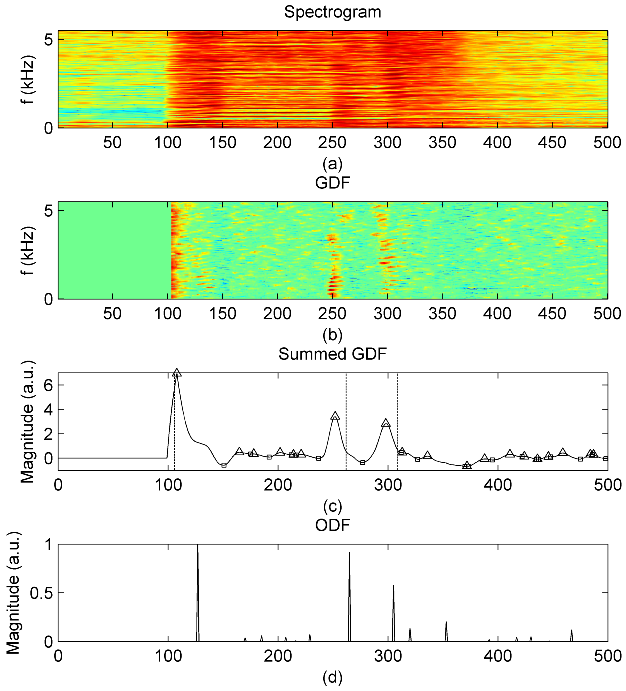

Figure 3. Illustration of PVGD method: (a) spectrogram, (b) masked GDF, (c) smoothed ODF with peak-valley marks and annotation, and (d) final ODF.

GDF, the negative derivative of phase (9) with respect to frequency, can be represented as:

$$\mathrm{GDF}(t,\omega) = \mathrm{Re}\left(-\frac{\mathrm{S}_x^{\mathcal{T}h}(t,\omega)}{\mathrm{S}_x^h(t,\omega)}\right) , \quad (10)$$

where $\omega = 2\pi f$ is the angular frequency and $\mathcal{T}(\cdot)$ is the operator such that $\mathcal{T}h(t) = t \cdot h(t)$. Please refer to [6, 7] for details of computing GDF. In this work, the group-delay function is computed by the Time-Frequency Toolbox (`http://tftb.nongnu.org/`). For brevity, we denote the discrete implementation of $\mathrm{S}_x^h(t,\omega)$ as $\mathrm{X}(n,k)$ and $\mathrm{GDF}(t,\omega)$ as $\mathrm{GD}(n,k)$ hereafter.

4.2 Proposed method

Onset events can be detected by zero-crossing or negative maximal difference (ΔGD) of GDFs [9, 10]. However, in noisy data, extracting informative zero-crossings is not an easy task, leading to false positives for ΔGD. From the discussion in Section 4.1, we found that a prominent peak (positive GDF) is usually followed by a prominent valley (negative GDF) — a property that has been mostly neglected in previous work. Our hypothesis is that considering both the peaks and valleys of GDF help differentiate the onset events from the noisy terms.

As exemplified in Fig. 3, the calculation of PVGD involves the following four consecutive steps:

1. **Preprocessing:** compute the raw spectrogram (Fig. 3 (a)) and GDF and then multiply the GDF by a binary mask (Fig. 3 (b)). The mask is computed from the spectrogram by setting the value of a time-frequency bin $\mathrm{X}(n,k)$ with increasing energy (w.r.t.

Table 1. F-scores and the corresponding thresholds of various onset detection methods; the symbol '*' denotes the proposed methods

Method		SF-based					GDF-based		WPD	CD
		$p = 2/3$*	$p = 1/2$*	$p = 1/3$*	SF ($p = 1$)	log-SF	PVGD*	Δ GD		
MAPS30	F-score	0.947	0.958	**0.962**	0.923	0.888	**0.963**	0.956	0.563	0.783
	δ	0.05	0.05	0.05	0.05	0.05	0.15	0.15	0.05	0.05
[9]	F-score	**0.754**	0.751	0.738	0.712	0.731	**0.742**	0.684	0.352	0.542
	δ	0.30	0.30	0.30	0.30	0.30	0.45	0.40	0.30	0.10

$\mathrm{X}(n-1, k)$) to 1 and a decreasing bin to 0. The bins with insignificant energy in the spectrogram (i.e., energy smaller than a thousandth of the maximum of the whole music piece) are also set to 0.

2. **Pooling:** sum up the masked GDF along the frequency axis and obtain a rough ODF (Fig. 3 (c)), which is subsequently smoothed to eliminate minor fluctuation terms (please see Section 5.2 for the smooth function).

3. **Peak-valley picking:** mark every peak-valley pair on the ODF and record the positions and values. Because an onset event includes a group-delay peak before onset perception and a valley after the onset, in principle every peak is followed by a valley. Fig. 3 (c) uses triangular marks to denote the peaks and rectangular marks for the valleys; we can see the peak at frame #108 and the valley at frame #151 forms a strong peak-valley pair, so do frames #252 and #275, and frames #298 and #313. Human annotations are depicted as dashed lines in Fig. 3 (c). We can observe that the ODF at frames #262 and #309 predict the onset positions accurately, whereas the ODF at frame #127 lags the onset position.

4. **Decision:** finally, we consider every middle point of a peak-valley pair as an onset and the magnitude difference between the peak and valley as the onset strength (Fig. 3 (d)).

5. EVALUATION

5.1 Dataset

We evaluated our methods on two datasets. The first dataset is a subset of the MIDI Aligned Piano Sounds (MAPS) database [13, 17]. We refer to the dataset as MAPS30, as it contains 30 piano pieces recorded by using an upright Yamaha Disklavier piano. The annotation data of MAPS includes the onset of every note even if they are played almost at the same time. To simplify the annotation data for an onset detection experiment, multiple onset events were regarded as a single event if they occur within 10 ms. This resulted in more than 10,000 onsets in total. Onset detection for this dataset is considered simpler as it has only one instrument.

The second dataset was compiled by Holzapfel *et al.* [9]. It is a more challenging dataset for onset detection as it encompasses 1,829 onset events for 12 different instruments classes including cello, clarinet, guitar, mixture, piano, saxophone, trumpet, violin, *kemençe*, *ney*, *ud*, and *tanbur*, with the last four being Turkish instruments.

5.2 Post-processing

One critical processing stage in onset detection task is the smoothing and peak picking procedures to exclude unwanted fluctuations [2, 3, 4, 5]. In this work, the following procedure was employed for all the detection methods except for PVGD. First, the raw ODF was smoothed by a Hanning window of length 10 (i.e., 29 ms; note this is different from the analysis window for computing STFT). Second, the ODF of a music piece was subtracted by its mean and divided by its standard deviation (z-scoring) for normalization. Third, an adaptive threshold was established by applying a median filter with length 100 (i.e., 290 ms) to the ODF function. Fourth, the ODF function was further subtracted by the adaptive threshold and then linearly normalized to the range $[0, 1]$. Onset events were detected using simple peak searching on the resulting curve.

PVGD does not require the aforementioned procedures, except for the first smoothing step, because it takes the middle point of a peak-valley pair as an onset directly. That is to say, PVGD is free from the extra parameters needed for adaptive thresholding.

5.3 Experiment setup

To reduce the computational cost, all audio files were down-sampled to 11,025 Hz first. Hanning window was adopted as the analysis window in computing STFT and GDF. The window length and the hop size were set to 1,024 samples and 32 samples (i.e., 2.9 ms), respectively. Note the hop size determines the finest resolution of onset detection. Valid peaks of the ODF were determined by thresholding.

Onsets were counted as correct detections when they are within a tolerance window of ± 50 ms around the onset annotation [9]. If two or more decisions were made within a tolerance window, only one decision was counted as a true positive, rendering others false positive. The accuracy for onset detection was evaluated in terms of F-score, the harmonic mean of precision and recall, following previous work [9, 10]. To investigate the effectiveness of the detection methods, we searched for the optimal F-score by exhausting the threshold δ from 0.05 to 0.95 with step size 0.05. The optimal F-scores were reported along with the corresponding thresholds.

Table 2. Performance of PVGD on the second dataset [9] using Hanning window with different lengths

window size	256	512	1024	2048	4096
F-score	0.658	0.706	0.742	**0.750**	0.733
δ	0.50	0.50	0.45	0.30	0.15

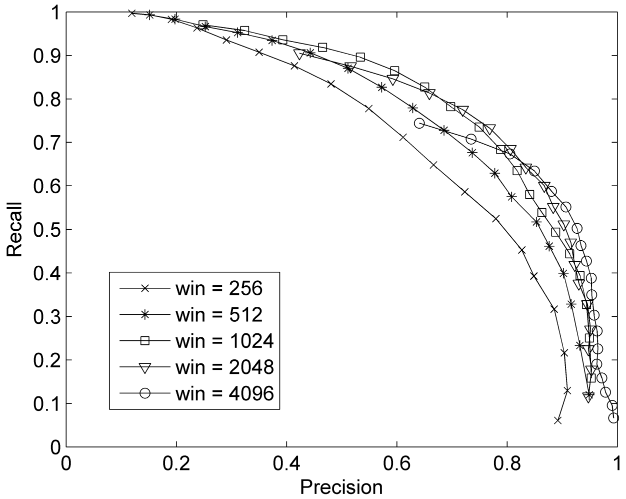

Figure 4. Precision-recall (P-R) curves of PVGD using Hanning window of various lengths for computing GDF.

5.4 Results

5.4.1 Overview

Table 1 lists the F-scores of all the methods discussed in Section 2. It can be found that, for the SF-based methods, setting $p < 1$ outperforms the conventional setting $p = 1$ for both datasets, with improvements ranging from about 2.5% to 4%. The optimal decision thresholds are not sensitive to the value of p; setting $p = 0.5$ seems to perform well. Moreover, using power-scale is found generally better than using logarithmic scale in both datasets. For GDF-based methods, PVGD performs comparably to ΔGD for MAPS30 but leads to significant improvement for the second dataset [9], improving the F-score from 0.684 to 0.742. The performance of PVGD for MAPS30 is not pronounced possibly because piano sounds inherently have sufficiently sharp attack envelope. We also note that WPD and CD are both inferior to the GD- and SF-based methods. The best two F-scores for the two datasets are indicated by bold fontface in Table 1. The proposed methods greatly outperform existing methods for the challenging dataset [9], and are also better than the results using only SF (0.741) and only GDF (0.737) reported in [9], respectively.

5.4.2 Effects of window sizes on GDF

After validating the effectiveness of the proposed methods, we move on to report the effect of the analysis window function. As discussed in Section 4, the performance of GDF is expected to be correlated with the length of the analysis window function, which influences the shape and length of the measured attack-decay slope. This is validated in Table 2, where we see great dependence of the F-scores and threshold values on the window size. Setting the window size to 2,048 slightly improves the F-score of PVGD to 0.750 for the second dataset. Window sizes of 256 (23.2 ms) or 512 (46.4 ms) are too short because the attack time duration may last up to over 300 ms for string and wind instruments [18]. On the other hand, although a long window with size 4,096 (0.37s) is enough to analyze soft attacks, such a window falls short of analyzing fast music, because it is easy to cover multiple onsets in a single frame.

Fig. 4 shows the P-R curves as we vary the threshold value from small values (higher recall) to large values (higher precision). We can see that long windows lead to relatively lower recall when the threshold value is small (left-hand-side), possibly implying that more groups of neighboring onset events are obscured into one (or even no) event in this case. In contrast, short windows lead to relatively lower precision (right-hand-side), possibly suggesting more prominent peaks are inaccurately located in this case. Better trade-off in precision and recall is obtained by using a moderate analysis window and a moderate threshold value.

6. DISCUSSION

Research has shown that fusing the decisions from different onset detection improves the overall performance remarkably [9]. Also, the uses of multi-resolution spectra, vibrato suppression or neural network are able to improve the robustness of the SF-based ODF [19]. In contrast to the best-performing methods which typically combined various approaches, the objective of this paper is to propose alternative methods by considering the nature of musical signals, such the dynamic range and the ADSR curve. Therefore, we opt for keeping the methodology simple to tackle the problem from a fundamental signal processing perspective.

7. CONCLUSIONS

In this paper, we have presented two novel methods that improve the robustness of onset detection against diverse musical dynamics and undesirable fluctuation in phase. Evaluation on two onset detection datasets with different number of instruments shows that the proposed methods are competitive alternative to existing ones. The proposed methods are conceptually simple and easy to be implemented. We conjecture that even better performance can be obtained by multi-band processing or decision fusion, as demonstrated in [10] and [9]. This is left as a subject of future study.

Acknowledgments

This work was supported by the Academia Sinica Career Development Award..

8. REFERENCES

[1] J. P. Bello, C. Duxbury, M. E. Davies, and M. B. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," *IEEE Signal Process Lett.*, vol. 11, no. 6, pp. 553–556, 2004.

[2] J. P. Bello, L. Daudet, S. A. Abdallah, C. Duxbury, M. E. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Speech Audio Process.*, vol. 13, no. 5-2, pp. 1035–1047, 2005.

[3] S. Dixon, "Onset detection revisited," in *Proc. Int. Conf. Digital Audio Effects*, 2006, pp. 133–137.

[4] F. K. Sebastian Böck and M. Schedl, "Evaluating the online capabilities of onset detection methods," in *Proc. Int. Soc. Music Information Retrieval Conf.*, 2012, pp. 49–54.

[5] C. Rosão, R. Ribeiro, and D. M. de Matos, "Influence of peak selection methods on onset detection," in *Proc. Int. Soc. Music Information Retrieval Conf.*, 2012, pp. 517–522.

[6] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the method of reassigment," *IEEE Trans. Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.

[7] S. Hainsworth, M. Macleod, S. W. Hainsworth, and M. D. Macleod, "Time frequency reassignment: A review and analysis," Cambridge University Engineering Department and Qinetiq, Tech. Rep., 2003.

[8] L. Su and Y.-H. Yang, "Sparse modeling for artist identification: Exploiting phase information and vocal separation," in *Proc. Int. Soc. Music Information Retrieval Conf.*, 2013.

[9] A. Holzapfel, Y. Stylianou, A. C. Gedik, and B. Bozkurt, "Three dimensions of pitched instrument onset detection," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 6, pp. 1517–1527, 2010.

[10] E. Benetos and Y. Stylianou, "Auditory spectrum-based pitched instrument onset detection," *IEEE Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 1968–1977, 2010.

[11] S. Böck and G. Widmer, "Local group delay based vibrato and tremolo suppression for onset detection."

[12] A. Röbel, "Onset detection in polyphonic signals by means of transient peak classification," *MIREX Online Proceedings (ISMIR 2005)*, 2005.

[13] V. Emiya, "Transcription automatique de la musique de piano," Ph.D. dissertation, TELECOM ParisTech, Paris, France, 2008, [Online] http://www.tsi.telecom-paristech.fr/aao/en/2010/07/08/maps-database-a-piano-database-for-multipitch-estimation-and-automatic-transcription-of-music/.

[14] S. Dixon, "Evaluation of the audio beat tracking system beatroot," *Journal of New Music Research*, vol. 36, no. 1, pp. 39–50, 2007.

[15] A. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 1999, pp. 115–118.

[16] J. O. Smith, *Introduction to digital filters: with audio applications.* W3K Publishing, 2007, vol. 2.

[17] C.-T. Lee, Y.-H. Yang, and H. H. Chen, "Multipitch estimation of piano music by exemplar-based sparse representation," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 608–618, 2012.

[18] S. McAdams, J. Beauchamp, and S. Meneguzzi, "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters," *J. Acoustical Society of America*, vol. 105 -103, pp. 882–897, 1999.

[19] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx). Maynooth, Ireland (Sept 2013)*, 2013.