IMPROVED ONSET DETECTION ALGORITHM BASED ON FRACTIONAL POWER ENVELOPE MATCH FILTER

Jian-Jiun Ding [1], *Chi-Jung Tseng* [2], *Che-Ming Hu* [3], *and Ta Hsien* [4]

[1,3,4] Graduate Institute of Communication Engineering in National Taiwan University
No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan
[2] Department of Applied English in Chihlee Institute of Technology
No. 313, Sec. 1, Wunhua Rd., Banciao District, New Taipei City, 220, Taiwan
TEL: +886-2-33669652, Fax: +886-2-33663662,
Email: djj@cc.ee.ntu.edu.tw, mindytseng@yahoo.com, truexray@hotmail.com, cocoid0@hotmail.com

ABSTRACT

In music and speech recognition, onset detection plays an important role for extracting the note of a music signal or the syllable of a speech signal. There are several existing onset detection algorithms, such as the difference of magnitude method, the short-term energy method, the surf method, and the high frequency content method. In this paper, we proposed an improved onset detection algorithm, which mainly applies the techniques of the fractional power amplitude, the envelop-matched filter, and other techniques to improve the accuracy and the efficiency of onset detection for humming signals. From simulations, our proposed method has both less computation time and higher accuracy than the existing methods. It also obviously increases the hit rate of the query-by-humming system.

1. INTRODUCTION

Onset detection is used for detecting the starts of a syllable or a music note. It plays an importance role for music signal processing, especially in the query-by-humming system. There are many existing onset detection algorithms [1-10]. The simplest one is the **difference of magnitude** method. It uses the difference of the envelope amplitudes of two consecutive time slots to detect the possible onset locations. Its process is:

(i) Determine the envelope amplitude:

$$A_k = \max(LPF\{x[n] \mid kn_0 \le n < (k+1)n_0\}), \quad (1)$$

where $x[n]$ is the input signal, n_0 is the width of time slot *LPF* is some lowpass filter.

(ii) $D_k = A_k - A_{k-1}$. (2)

(iii) If D_k > threshold, then kn_0 is recognized as a location of the onset.

The method is simple and intuitive. However, its performance is highly affected by the background noise, since it just considers the difference of the values of A_k. A humming signal usually has large background noise. Moreover, if the amplitude does not increase abruptly at the beginning of the note, than the onset cannot be detected from the difference in (2). Furthermore, if the noise is wider than the time slot, which often happens, then using the *LPF* in each time slot as in (1) may not remove the noise. It also increases the computation time.

Another onset detection method is the **short-term energy method**. It is also easy for implementation. Its process is:

(i) $$E_k = \sum_{n=kn_0}^{(k+1)n_0-1} x^2[n], \quad (3)$$

(ii) $$D_k = E_k - E_{k-1}, \quad (4)$$

(iii) If D_k > threshold, then kn_0 is recognized as a location of the onset.

The process is similar to that of the difference of magnitude method. The difference is that the energy is used as the feature due to human perception. The method is more effective for the psychoacoustic onset detection. However, it is also sensitive to noise.

In [9], Pauws proposed another onset detection method, i.e., the **surf method**. He used the slope of the envelope to detect the onset, where the slope is determined by the quadratic polynomial approximation. The detail of the process is:

(i) Apply (1) to find the envelope amplitude A_k in each time slot.

(ii) Approximate A_m ($m = k–2 \sim k+2$) by a second order polynomial $p[m] = a_k + b_k(m–k) + c_k(m–k)^2$. The coefficients a_k, b_k, and c_k can be solved by the least mean square error solution. For example, b_k can be determined from:

$$b_k = \sum_{\tau=-2}^{2} A_{k+\tau}\tau \Big/ \sum_{\tau=-2}^{2} \tau^2 . \quad (5)$$

(iii) b_k has the physical meaning of slope. If b_k > threshold, then we can conclude that kn_0 is the location of the onset.

The surf method is more precise and has higher robust to noise. However, in addition to (1), the method requires extra computation time to determine the slope b_k in (5) for each k.

In [7], Maris and Bateman proposed the **high-frequency content (HFC) method** for onset detection. It is based on the concept that, at the location of the onset, the signal usually has more high frequency components. Its process is :

(i) First, perform the DFT for the k^{th} time slot:

$$X_k[m] = \sum_{q=0}^{n_0-1} x[kn_0+q]e^{-j\frac{2\pi}{n_0}mq} . \quad (6)$$

(ii) Then, calculate the total energy E_k and the high frequency energy H_k in the k^{th} time slot:

$$E_k = \sum_{m=0}^{n_0-1} |X[m]|^2 , \quad H_k = \sum_{m=0}^{n_0-1} |X[m]|^2 \, w[m] ,$$

where the weighting function $w[m] \approx 1$ when m is near to $n_0/2$ and $w[m] \approx 0$ when m is near to 0 or n_0.

(iii) $$DF_k = \frac{H_k}{H_{k-1}} \cdot \frac{H_k}{E_k}. \quad (7)$$

(iv) If $DF_k >$ threshold, then kn_0 is recognized as a location of the onset.

In Step 1, the DFT can also be replaced by the wavelet transform [10]. The HFC method is very reasonable. However, its performance is also affected by the background noise.

In this paper, we propose a new onset detection algorithm. We apply the techniques of the fractional power amplitude, the envelope match filter, and other techniques. See Section 2. Then, in Section 3, we show, that with the proposed algorithm, both the accuracy and the computation efficiency of onset detection are significantly improved. It also increases the hit rate of the query by humming system.

2. PROPOSED ALGORITHM FOR ONSET DETECTION

A good onset detection algorithm should achieve the following two goals:

(a) Accuracy: For a music signal, if there is an onset at n_1, then after applying the onset detection algorithm, we should find that there is an onset around the location n_1 (i.e., *higher true positive rate)*. Moreover, if we find an onset at the location n_2 by the onset detection algorithm, then the original music signal should indeed have an onset around n_2 (i.e., *lower false positive rate*).

(b) Efficiency: The computation time should be smaller.

To achieve the first goal, the effect of the background noise should be reduced. Moreover, we should avoid the misjudgement caused from the trill, the warble tone, and the end tone of the music signal. Based on this requirement, we proposed the techniques of the "fractional power amplitude", the "envelope match filter", and "background noise compensation" to improve the accuracy of onset detection.

To achieve the goal of efficiency, we think that the process of performing the lowpass filter (or the DFT) in each time slot as in (1) and (6) can be avoided. We suggest that one only has to find the value of

$$A_k = \max(\{|x[n]| \mid kn_0 \le n < (k+1)n_0\}) \quad (8)$$

for each time slot instead of (1), (3), and (6). With the modification, the computation time can be much reduced.

It seems that the effect of the noise cannot be reduced without using the lowpass filter in each time slot. However, in fact, we find that performing the match filter on A_k in (8) has higher ability to reduce the effect of the noise. Furthermore, the thrill, the warble tone and the end tone of the music signal, which may lead to the misjudgement of the onset, often have the width larger than that of the time slot. Therefore, it is more reasonable to perform the filter on A_k instead of $x[n]$ ($kn_0 \le n < (k+1)n_0$), as in Fig. 1.

The details of the applied techniques are described in the following four subsections.

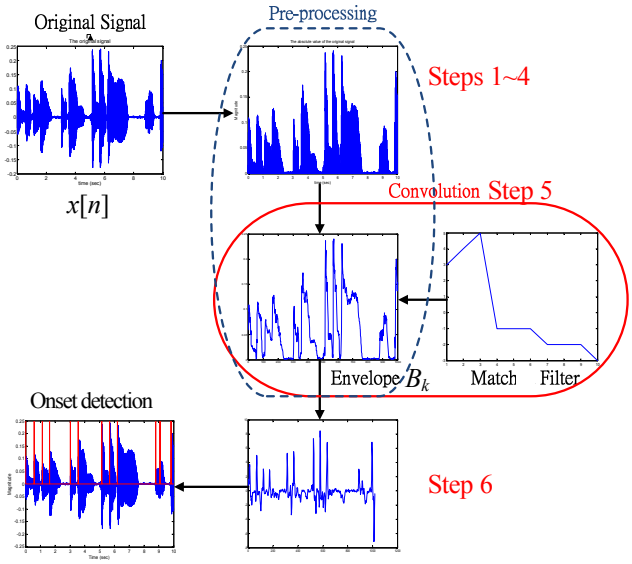

Figure 1 – A simple flowchart of our proposed onset detection algorithm. The detail of the process is shown in subsection 2-4.

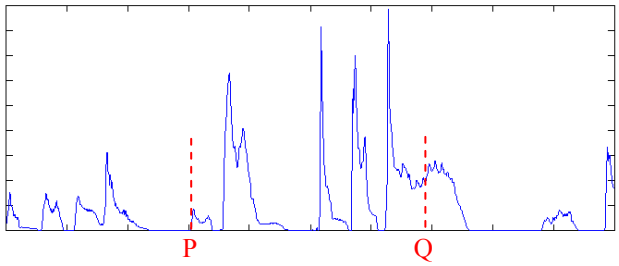

Figure 2 – The envelope amplitude of a music signal. There should be an onset at the location P but no onset at Q.

2.1 Fractional Power Amplitude

The conventional onset detection algorithms usually use the variation of the envelope amplitude or its energy to judge whether there is an onset. However, in practice, observing the variation of the envelope amplitude directly will often lead to misjudgement.

For example, for a music signal whose envelope amplitude is as in Fig. 2, if we use (2) to calculate the difference of the envelope amplitude, then we find that value of D_k at the location Q is a little larger than that at the location P. However, in fact, there should be an onset at P but no onset at Q. The energy of the music signal increases at the location Q is due to the thrill of the music signal, not due to the new note. By contrast, the value of D_k is small at P is due to that the voice is small around the location. In fact, there is indeed an onset at P.

Therefore, we suggest that it is proper to take the fractional power for the envelope amplitude. That is,

$$B_k = A_k^{\lambda}, \quad \text{where } 0 < \lambda < 1 \quad (9)$$

and A_k is the envelope amplitude. Suppose that $A_k = 0.1$, $A_{k-1} = 0$, $A_h = 0.52$, and $A_{h-1} = 0.4$. If we use (2) to compute the difference, then $D_h = 0.12 > D_k = 0.1$. By contrast, if we take the fractional power as in (9) and choose $\lambda = 0.7$, then

$$B_k - B_{k-1} = 0.1995 > B_h - B_{h-1} = 0.1062 .$$

Therefore, with the fractional power amplitude in (9), the misjudgement caused from the thrill of the music signal can be avoided.

2.2 Envelope Match Filter

As the description in Section 1, many exiting onset detection algorithms use the difference of some quantities of two adjacent time slots to find the onset, as in (2) and (4). Note that the difference operations in (2) and (4) are equivalent to the convolution of A_k with the following filter

$$[1, -1]. \tag{10}$$

The HFC method in (7) is hard to be expressed in the convolution form. However, it essentially observes the high frequency energy variation of two adjacent time slots. The surf method is different from other methods and use five adjacent time slots to determine the "slope", as in (5). Note that, from (5), the surf method is in fact equivalent to the convolution of A_k with the following filter:

$$[2,\ 1, 0, -1, -2]/10. \tag{11}$$

In this paper, we think that applying the match filter instead of (10) and (11) is a more reasonable way for onset detection. The concept of the match filter is to use the time reversal of the pattern as the filter to find the desired object. However, since a music signal note usually has no fixed frequency and no fixed shape, applying the match filter for the music signal directly may not have good performance. Instead, we suggest that it is proper to apply the match filter to the fractional scaling of the envelope amplitude in (9).

From our statistics, we find that, if we choose the width of the time slot as 0.01 Sec, then the average of the fractional scaled envelope amplitude in the onset region is near to:

$$\eta\cdot[\text{-2,-2, -2, -2, -2, -2, -1, -1, 4, 4, 3, 3}],$$

where η is some constant. Therefore, the envelope match filter we choose for onset detection is

$$f[n] = [3, 3, 4, 4, \text{-1, -1, -2, -2, -2, -2, -2, -2}]. \tag{12}$$

Its length is 12 and the waveform is plotted in Fig. 3.

The performance of the proposed method will be better than those of the conventional methods that consider only the difference of the quantities of two or five adjacent time slots. Moreover, compared with the filters in (10) and (11), the proposed match filter in (12) is more similar to the time reversal of the envelope of the music signal near the onset regions. Therefore, using the proposed envelope match filter can improve the performance of onset detection.

Furthermore, the envelope match filter also provides an effective and efficient way to remove the noise. Many conventional algorithms apply the lowpass filter for each time slot, as in (1). It is a good way to remove the noise if the noise is narrower than the time slot. However, to increase the resolution of the onset detection, the width of the time slot is always chosen as a small value (about 0.01 Sec). For a music signal, the noise caused from thrill or the warble tone usually wider than 0.01 Sec. Therefore, using the LPF in (1) may not remove the noise effectively. It is more reasonable to perform the filter on B_k in (9). Therefore, applying the envelope match filter in (12) to the scaled envelope amplitude B_k in fact has higher ability to remove the noise. It also much reduces the computation time.

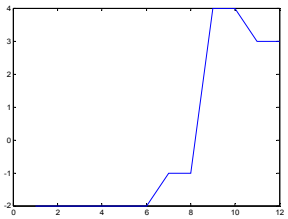

Figure 3 –Proposed envelope match filter.

2.3 Other Improvement Techniques

Furthermore, we also apply some minor techniques to improve the performance of onset detection. To further remove the effect of the background noise, we can apply the following operation to perform the background noise compensation:

$$A_{out,k} = \max\left(A_{in,k} - \rho, 0\right). \tag{13}$$

The value of ρ can be chosen according to the magnitude of the background noise. For the *.wav file generated from the microphone, we can choose $\rho = 0.02$.

Moreover, to obtain a more objective onset detection result, it is proper to perform the normalization as follows before applying the envelope match filter:

$$A_{out,k} = \frac{A_{in,k}}{0.2 + 0.1\cdot E}, \tag{14}$$

where E is the mean of the mean of $A_{in,k}$.

2.4 Process of the Proposed Onset Detection Algorithm

The proposed onset detection algorithm is summarized as:

(Step 1) First, use (8) to find the envelope amplitude for each time slot.

(Step 2) Then, use (13) to reduce the effect of the background noise.

(Step 3) Use (14) to normalize the envelope amplitude.

(Step 4) Take the fractional power of the envelope amplitude by (9).

(Step 5) Then, perform the convolution of B_k (obtained in (9)) and the match filter in (12).

$$C_k = \sum_{\tau=0}^{11} B_{k-\tau} f[\tau]. \tag{15}$$

(Step 6) If C_{k+3} > threshold, then kn_0 is recognized as a location of the onset.

3. SIMULATIONS

We show an example of using our proposed algorithm to detect the onset of a humming signal in Fig. 4. The results of each step are shown in Figs. 4(b)-4(e). From the result in Fig. 4(e), we can see that all onsets of the humming signal are detected successfully.

We perform another simulation in Fig. 5, which uses the difference of magnitude method, the short-term energy method, the surf method, the HFC method (these four methods were described in Section 1), and the proposed method to detect the onset of the humming signal in Fig. 5(a). The onset detection results for each method are shown Figs. 5(b)-5(g). The original humming signal actually has 12 onsets.

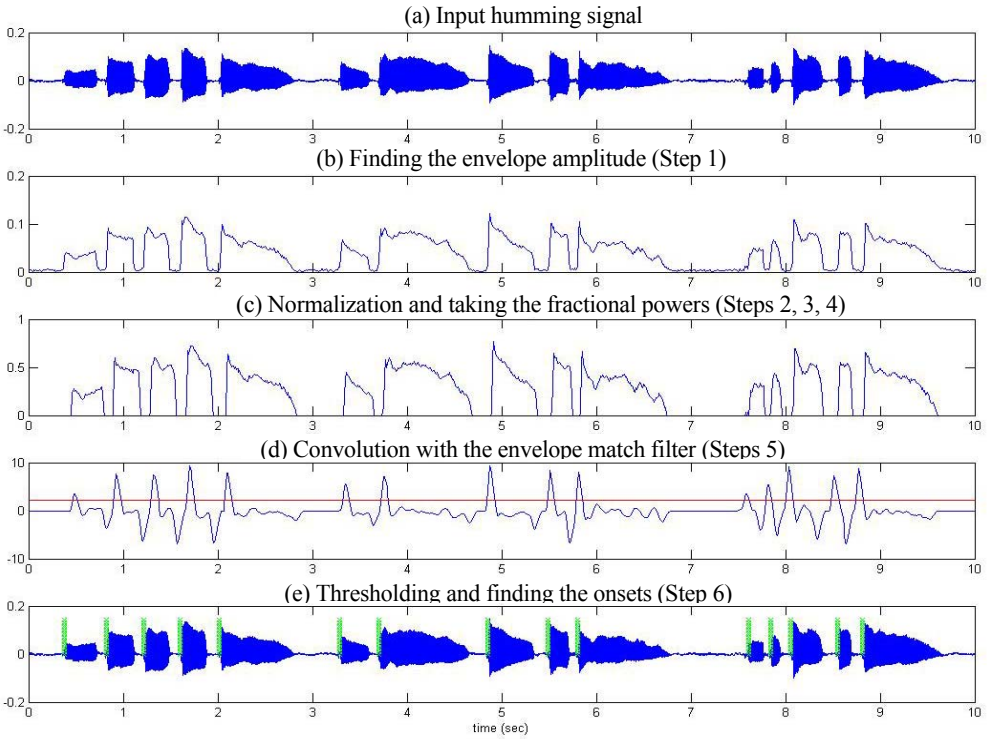

Figure 4 – An example of onset detection using our proposed algorithm.

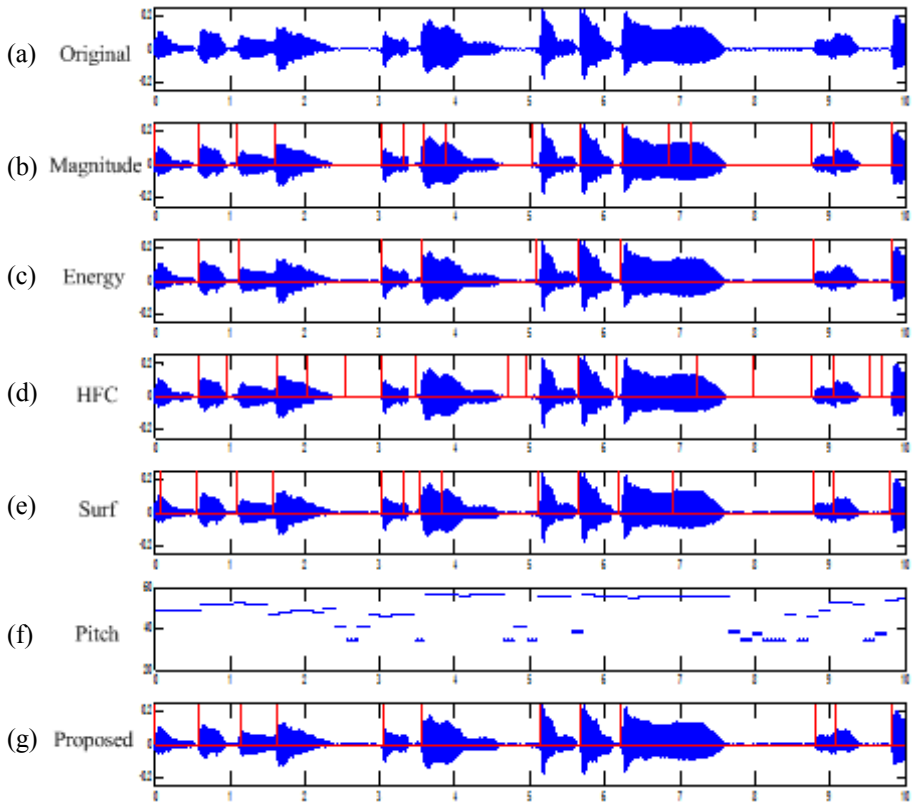

Figure 5 – The onset detection results using the existing and our proposed algorithms. The tested music signal actually has 12 notes.

From Fig. 5(g), we can see that, using the proposed algorithm, the onsets of the original humming signal are detected perfectly. When using other onset detection methods, the results are affected by the thrill and the rising of the end tone. Using the proposed method can avoid these problems.

Then, in Table I, we perform another simulation. We use the database that contains 70 humming signals as the input [11]. 26 of the humming signals are 20 second length and others are 15 second length. We use the true positive rate and the false positive rate to measure the performance, where

TABLE I. The results of onset detection for the database that contains 70 humming signals. Each humming signal is 15 or 20 second length.

Methods	True positive rate	False positive rate	Total computation time
Difference of magnitude	90%	16%	3.43 Sec
Short-term energy	78%	26%	0.65 Sec
HFC method	93%	5%	3.61 Sec
Surf method	97%	12%	2.52 Sec
Proposed method	**99%**	**2%**	**0.23 Sec**

TABLE II. The hit rates of the query-by-humming systems that use a variety of onset detection algorithms (including the proposed one). The melody matching method is fixed to the DP method in [12].

	Difference of magnitude	Short-term energy	HFC method	Surf method	**Proposed method**
hit rate	80.0%	64.3%	87.1%	88.6%	**97.1%**

$$true\ positive\ rate = \frac{TP}{TP+FN},$$

$$false\ positive\ rate = \frac{FP}{TP+FP}, \quad (16)$$

TP: the number of the onsets of the original signal that are detected by the onset detection algorithm,

FN: the number of the onsets of the original signal that are not detected by the onset detection algorithm,

FP: the number of the detected onset that are in fact not the onset of the original signal.

From Table I, we can see that the proposed method has both higher true positive rate and lower false positive rate than other methods. Furthermore, the proposed algorithm also has much smaller computation time than other methods, since in our algorithm the computation of the lowpass filter, the energy sum, or the DFT for each time slot is saved.

In Table II, we show the performances of the query-by-humming systems that use the existing and the proposed onset detection algorithms together with the melody matching method of dynamic programming (DP) [12]. The tested data are the same as that in Table I. From the hit rates shown in Table II, the proposed onset detection algorithm can significantly increase the hit rate of the query-by-humming system and improve the accuracy of music retrieval.

4. CONCLUSION

In this paper, we proposed an improved onset detection algorithm. With the techniques of the fractional power amplitude, the balance of the background noise, envelope magnitude normalization, and the envelope match filter, the performance of onset detection can be much improved. Moreover, since the lowpass filter and the DFT in each time slot are avoided, the computation time is also smaller. The proposed onset detection algorithm is helpful for improving the efficiency and the accuracy of the query-by-humming system.

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 Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035-1047, Sept. 2005.

[2] S. Hainsworth and M. MacLeod, "Onset detection in musical audio signals," *Proc. Int. Computer Music Conference*, pp. 163-166, 2003.

[3] J. P. Bello and M Sandler, "Phase-based note onset detection for music signals," *ICASSP*, vol. 5, pp. 441-444, Apr. 2003.

[4] A. M. Barbancho, A. Jurado, and I. Barbancho, "Identification of rhythm and sound in polyphonic piano recordings," *Proc. of Forum on Acousticum*, Sevilla, 2002.

[5] C. Duxbury, M. Sandler, and M. Davies, "A hybrid approach to musical note onset detection," *DAFx-02*, pp. 33–38, Sept. 2002.

[6] J. Foote, "Automatic audio segmentation using a measure of audio novelty," *ICME*, vol. 1, pp. 452–455, July 2000.

[7] P. Masri and A. Bateman, "Improved modelling of attack transients in music analysis-resynthesis," *Proceedings of the International Computer Music Conference*, Hong-Kong, pp. 100-104, Aug. 1996.

[8] J. P. Bello, G. Monti, and M. Sandler, "Techniques for automatic music transcription," *ISMIR*, Polymouth, Massachusetts, 2000.

[9] S. Pauws, "CubyHum: A fully operational query by humming system," *ISMIR*, pp. 187-196, 2002.

[10] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293-302, July 2002.

[11] http://djj.ee.ntu.edu.tw/TestSongs.zip

[12] Y. Zhu and M. S. Kankanhalli, "A robust music retrieval method for query-by-humming," *International Conference on Information Technology: Research and Education,* Singapore, pp. 89-93, Aug. 2003.

[13] P. Brossier, *Automatic Annotation of Musical Audio for Interactive Applications*, Ph.D. Thesis, Queen Mary University of London, UK, August 2006.

[14] A. Lacoste and D. Eck, "Onset detection with artificial neural networks for MIREX 2005," *MIREX*, London, UK, 2005.

[15] A. Robel, "Onset detection in polyphonic signals by means of transient peak classification," *MIREX*, Victoria, Canada, 2006.

[16] W. C. Lee, Y. Shiu, and C. C. J. Kuo, "Musical onset detection with linear prediction and joint features," *MIREX*, Vienna, Austria, 2007.

[17] A. Robel, "Onset detection by means of transient peak classification in harmonic bands," *MIREX*, Kobe, Japan, 2009.

[18] F. Eyben, S. Bock, B, Schuller, and A. Graves, "Universal onset detection with bidirectional long short-term memory neural networks," *ISMIR*, pp. 589-594, 2010.