PITCH POWER FLUX METHODS EMPHASIS ON STRING INSTRUMENT FAMILY FOR MIREX 2012 ONSET DETECTION

Chang-Hyun Kim
KAIST
Dep. Bio and Brain Engineering
flipflop98@gmail.com

Won-il Chang, Sang-Hoon Oh
KAIST, Mokwon Univ.
Dep. Bio and Brain Engineering, Dep. Information & Communication Eng.
shoh@mokwon.ac.kr, chang.wonil@gmail.com

Soo-Young Lee
KAIST
Dep. Electrical Engineering
sy-lee@kaist.ac.kr

ABSTRACT

This extended abstract proposes a new method for onset detection using harmonic information on each note. Especially, this method is optimized on string instrument family irrespective of articulation. This method outperforms about 31 and 24% in F-measure than existing algorithms such as spectral flux or spectral difference.

1. INTRODUCTION

Onset detection is important preprocessing step for audio signal processing. In the automatic transcription system, many previous researches about this problem based on learning [1] and signal processing [2][3][4][5] has been reported. Masri[5] has proposed the spectrum based onset detection functions such as dissimilar function. Juan[2] has proved that many spectrum power flux based onset detection algorithms have robust and fast performance. Recently, Sebastian Bock's BLSTM-RNN[1] algorithm based on three dimensional recurrent neural networks learning algorithm has become the state of art performance in MIREX 2011 onset detection. However, this algorithm is time expensive owing to the multi-layer neural network learning. Therefore, it is required to build time inexpensive and robust working algorithms for onset detection. In Chapter 2, we will propose two onset detection functions based on music instrument acoustics. In Chapter 3, four parameters are introduced for improving the onset detection performance. Test environment and evaluation methods are explained in Chapter 4 and we will show the test results followed by discussion.

2. ONSET DETECTION FUNCTIONS

We propose two onset detection functions based on note-scale filter bank power spectrum.

2.1 Note-Scale Filter Bank

As Figure 1 shows, the note-scale spectrum is composed of five blocks; HPF (High Pass Filtering), Windowing, |STFT|2 (Power Spectrum), Note-Scale Filter Banks, and Log. Firstly, the high pass filtering is the first order FIR filter with [1 -0.97] coefficients, which pre-emphasize the high frequency components. We do this process for compensating for the weak high frequency power in the note spectrum. The windowing is performed for framing the time domain signal and do the short time fourier transform square for building power spectrum. Then, we weight and sum with the triangular note scale filter bank on this power spectrum. Lastly, we convert the note-scale filter bank output with log, which is an inevitably required process to discriminate more between the weak and strong power component in the log spectrum.

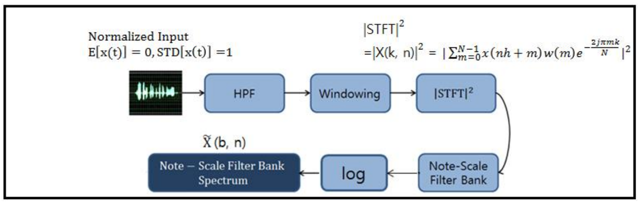

Figure 1. Note-Scale Filter Bank Spectrum Extraction Procedure

In Figure 2, the Note-scale filter banks have center frequencies as the musical note fundamental frequency, F0s We designed each of the previous filter banks' center and end frequencies overlapping with the following filter banks' the start and center frequencies separately. The height has triangular shape with decreasing amplitude. It preserves the same energy on each filter bank. Lastly, we designed 103 filters, covering fully the weaker and higher power harmonics in music instruments.

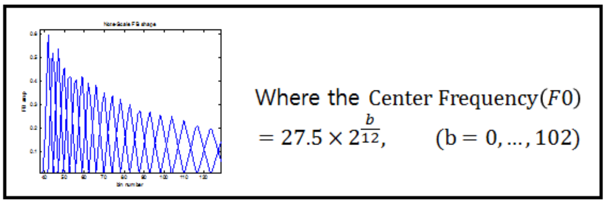

Figure 2. Note-Scale Filter Banks Design

2.2 Pitch Power Flux (PPF)

Pitch Power Flux is the colume-wise sum of the two consecutive note-scale power spectrum difference on each pitch. To detect only onset points except offsets, we passed half-wave rectifier to the spectrum difference as formula 1 defines.

$$PPF(F_0, n-1) = \sum_{b \in B} H(\widetilde{X}(b,n) - \widetilde{X}(b,n-1)) \quad (1)$$

,where H represents half-wave rectifier.

For PPF and NPPF methods, we used the following (FIGURE 3) operational algorithm. Firstly, the silence pitch in a frame is ignored by pre-filtering with delta value. If the pitch is over delta values, we put it as a onset candidate frame. Among all pitches, we select the biggest value and check the validity as a real onset through the comparison with the common threshold value. We choose the onset frame following this framework.

```
for n=2:T
  for F0=1:88
      if X̃'(F0, n) ≧ δ
            active_F0 = [active_F0 F0];
      end
  end
  PPF or NPPF = (max(PPF(1:88 ,n-1) or NPPF(1:88 ,n-1))>Threshold)
end

,where X̃'(F0, n)=∑_{b∈B} X̃(b, n)/{number of B set elements}
B = [F0 F0+12 F0 + 19 F0 + 24 F0 + 28 F0 + 31 F0 + 34 F0 + 36]
and 'F0' is the pitch number.
```

FIGURE 3. PPF and NPPF Operation Pseudo Code

2.3 NPPF (Normalized Pitch Power Flux)

For increasing the musical harmonic information in the feature, we add the first eight integer multiple frequency note filter bank bins on each pitch allowing repetition as formula 2 shows.

$$NPPF(F_0, n-1) = \frac{\sum_{b \in B} H(\widetilde{X}(b,n) - \widetilde{X}(b,n-1))}{\sum_{b \in B} \widetilde{X}(b,n)} \quad (2)$$

,where H represents half-wave rectifier.

In formula 1 and 2, the P is the pitch number from 1 to 88 representing A0 and C8 notes in music. B is the note filter bank bin set of the first eight integer multiple fundamental frequencies. (Table 1)

Octave #	Note #	Hz Ratio (Frequency/F0)
1	1 (F0)	1
2	13	2
3	20	2.996614≈3
4	25	4
5	29	5.039684≈5
6	32	5.993228≈6
7	35	7.12719≈7
8	37	8

Table 1. Note Number Frequency Ratio (Frequency and F0s are calculated through Center Frequency Equation)

In the high frequency notes, we add up to 103^{th} note filter bank numbers.

3. PARAMETER SET

There are four parameters to improve the performance.

3.1 Delta Value

For pre-filtering out the silence in one time frame, we checked the note filter bank power amplitude on each note fundamental as Figure 3 shows.

3.2 Threshold Value

The threshold value separates onset detection function's peak values into two groups; valid onset points and invalid onset points.

3.3 Time Delay Offset

On the time axis, the true positive points are detected points in the tolerance interval. We draw a histogram with the distribution of these true positive points and find the 50% existence point in the histogram. Then, we moved this point back and forth to locate on the same spot as the ground truth onset point. This time difference is called time delay offset.

3.4 Merge detected notes(Post Processing Process)

At the post processing, we merged onset detection functions' peaks, which are located in the pre-defined interval. Here, the shortest length note in the music is very important information for selecting the merging time.

4. TEST

4.1 TEST Procedure

Step 1.	Input Mixture Data Normalization
Step 2.	Transformation into Note-Scale Filter Bank
Step 3.	Calculate Onset Detection Function
Step 4.	Peak detection
Step 5.	Thresh-holding
Step 6.	Post Processing
Step 7.	Evaluate (with tolerance length, ±25msec)

Table 2. Bello Pitched Non-Percussive (PNP) set Test Procedure

We checked our algorithms with bello set pitched non-percussive (PNP). Test procedure is composed of seven steps. We preprocessed the 1 file, 93 onsets in a chopped length music with mean zero and standard deviation one and transformed the time signal into note-scale filter bank output. We calculate PPF and NPPF onset detection value sequences and do the peak detection and lastly do the peak picking process through thresh-holding with the pre-optimized value. Comparing to and Bello's PP set test, this experiment has two times bigger threshold value. We evaluated the tolerance interval length with ±25msec.

4.2 Evaluation Method

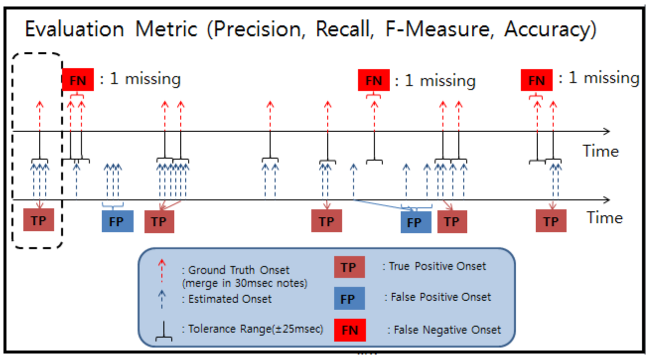

Figure 4. Evaluation Example

From the Figure 4, the first three detected onsets are true-positive points inside the tolerance range. In the second group, only one onset is detected and one is missing, which makes false negative point. All other remains in the third group points are false positives.

5. RESULTS

5.1 Experiment Result

As Table 3 shows, PPF has the best Precision, Recall and F-measure values 0.9195, 0.8696 and 0.8939 among three other onset detection functions. Comparing to the conventional method such as SF and SD, the NPH and NHPF has better performance in general.

	Precision	Recall	F-measure
SF	0.8542	0.4457	0.5857
SD	0.7051	0.6044	0.6509
PPF	**0.9195**	**0.8696**	**0.8939**
NPPF	0.6667	0.6739	0.6703

Table 3. Bello set PNP test result

(SF and SD represent spectral flux and spectral difference each.)

6. DISCUSSION

If we change the two test processes; peak detection and thresh-holding, we will miss some peaks, which have only peak point over the thresh-holding line.

7. REFERENCES

[1] F. Eyben, S. Bock, B. Schuller, A. Graves: "Universal Onset Detection with Bidirectional Long-Short-Term Memory Neural Networks," *Proceedings of the International Symposium on Music Information Retrieval*, pp. 589–594, 2010.

[2] 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. On Speech and Audio Processing*, Vol. 13, No. 5, pp. 1035–1047, Sept. 2005.

[3] S. Dixon. "Onset detection revisited," *Proceedings of DAFx-06*, pp. 133-137, Sept. 2006.

[4] A. Robel. "Onset Detection in Polyphonic Signals by means of Transient Peak Classification," *Proceedings of the International Symposium on Music Information Retrieval*, MIREX Onset Detection webpage, 2006.

[5] P. Masri and A. Bateman. "Improved modeling of attack transients in music analysis-resynthesis," *Proceedings of the International Computer Music Conference*(ICMC), pp100-103, 1996.

[6] L. Cohen: *Time-frequency analysis*, Signal Processing Series, Prentice Hall, 1995