ONSET DETECTION USING COMB FILTERS

Mikel Gainza, Eugene Coyle

Digital Media Centre
Dublin Institute of Technology
Dublin, IRELAND
[mikel.gainza,eugene.coyle]@dit.ie

Bob Lawlor

Department of Electronic Engineering
National University of Ireland
Maynooth, IRELAND
rlawlor@eeng.may.ie

ABSTRACT

A technique for detecting note onsets using FIR comb filters which have different filter delays is presented. The proposed onset detector focuses on the inharmonic characteristics of the onset component and the energy increases of the signal. Both properties are combined by utilizing FIR comb filters on a frame by frame basis in order to obtain an onset detection function, which is suitable for detecting slow onsets. The proposed approach improves upon existing methods in terms of the percentage of correct detections in signals containing slow onsets.

1. INTRODUCTION

A musical onset is defined as the precise time when a new note is produced by an instrument. The onset of a note is very important in instrument recognition, as the timbre of a note with its onset removed can be very difficult to recognize. Masri [1] states that in traditional instruments, an onset is the stage during which resonances are built up, before the steady state of the signal. Other applications use separate onset detectors in their systems, for example in rhythm and beat tracking systems [2], music transcriptors [3-7], time stretching [8], or music instrument separators [9].

The onset detectors encounter problems in notes that fade-in; in ornamentations such as grace notes, trills or cuts and strikes in traditional Irish music; and in fast passages such as arpeggios, or legatos. Also, the physical attributes of the instruments and recording environments can produce artifacts, resulting in detection of spurious onsets. Amplitude and frequency modulations that take place in the steady part of the signal can also result in inaccurate detections.

Section 2 provides an overview of the existing approaches that have dealt with the onset detection problem. Section 3 focuses on comb filter theory. In section 4 the proposed onset detector method is presented. Some results which validate the approach are shown in section 5 and finally, some conclusions and further work are discussed in section 6.

2. EXISTING APPROACHES

There are many different classifications of onsets. However, the two most common are:

A fast onset, which is a short duration of the signal with an abrupt change in the energy profile, appearing as a wide band noise burst in the spectrogram (see Fig.1). This change manifests itself particularly in the high frequencies and is typical in percussive instruments.

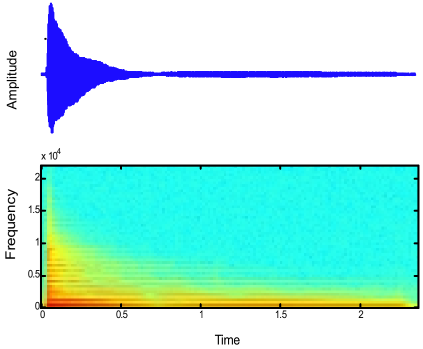

Figure 1: *Spectrogram of Piano playing F5*

Slow onsets which typically occur in wind instruments like the flute or the tin whistle, are more difficult to detect. In this case, the onset takes a much longer time to reach the maximum onset amplitude value and has no noticeable change in the high frequencies.

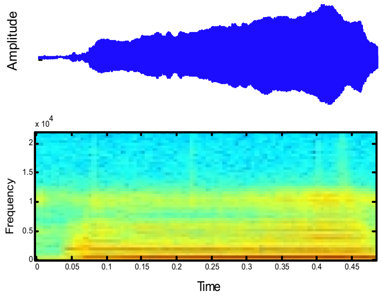

Figure 2*: Spectrogram of a tin whistle playing F#5*

A significant amount of research on onset detection and analysis has been undertaken. However, accurate detection of slow onsets remains a significant problem.

Energy based systems have been widely utilized for detecting onsets. Early work which dealt with the problem analyzed the amplitude envelope of the entire input signal for the purpose of onset detection [10]. To obtain information on specific frequency regions where the onset occurs, Bilmes suggested a multi band approach which computes the short time energy of a high frequency band using a sliding window [11], and Masri in [1], gives more weight to the high frequency content (HFC) of the signal. In [2], Scheirer presents a system for estimating the beat and tempo of acoustic signals requiring onset detection. A

0-7803-9154-3/05/$20.00 ©2005 IEEE

filterbank divides the incoming signal into six frequency bands, each one covering one octave, the amplitude envelope is then extracted, and the peaks are then detected in every band. Klapuri in [12], developed an onset detector system based on Scheirer's model. He utilizes a bank of 21 non-overlapping filters covering the critical bands of the human auditory system, and incorporates Moore's psycoacoustic loudness perception model [13] into his system. Klapuri obtains the loudness of every band peak, then combines all peaks together sorted in time, and calculates a new peak value for every onset candidate by summing the peak values within a 50 ms time window centered at the onset candidate. All the above approaches behave well for sharp onsets and for signals with a rich harmonic content.

In order to obtain more accurate results on detecting slow onsets, an approach that customizes an energy based onset detector system according to the characteristics of the Irish tin whistle is presented in [6]. A filterbank splits the signal into one band per note that the tin whistle can play, and different band thresholds are set according to expected note blowing pressure.

An alternative to energy based onset detection is proposed in [14], by calculating the frame by frame distribution of the differential angles by using phase vocoder theory. This approach is more sensitive than standard energy based approaches for analyzing soft onsets.

A combination of energy and phase information to favor the detection of slow and sharp onsets is presented in [15].

3. FIR COMB FILTER THEORY

By using FIR comb filters, the comb spectral shape can be obtained by summing an input signal x[n] with a delayed version of the same signal [16]. The FIR comb filter transfer function and the difference equation are represented as follows:

$$y[n]=x[n]+g*x[n-D] \tag{1}$$

$$H(z)=1+g*z^{-D} \tag{2}$$

where g is a factor which scales the gain of the filter between 1+g and 1-g, and D is the delay in samples.

The comb effect results from phase cancellation and summation between the delayed and undelayed signal. This can be appreciated in Figure 3 where the magnitude responses of a filter with g = 1, sampling rate fs = 44100 Hz and a delay D = 16 is depicted.

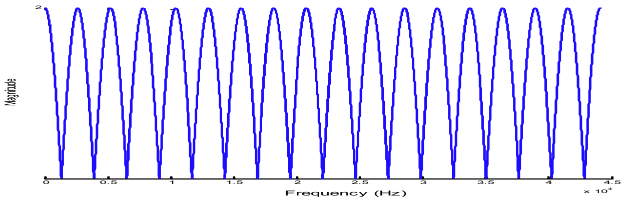

Figure 3: FIR comb filter magnitude response using g=1, D=16 and fs=44100 Hz

From Figure 3, it is apparent that at frequencies:

$$n*\frac{f_s}{D}<=f_s \tag{3}$$

where n is an integer, the delay D causes a 360 degree shift between the original and delayed signal causing addition, which produces peaks in the filter magnitude response at frequencies denoted by equation 2.

Thus, the energy of the signal $x[n]$ is doubled only if the peaks of the signal coincide with the peaks of the FIR comb filter. This will only occur for a given delay D and its integer multiples.

4. PROPOSED APPROACH

A technique for detecting onsets by using comb filters is proposed. In Figure 4, a block diagram illustrating the different components of the system is depicted.

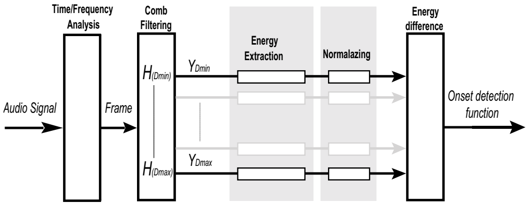

Figure 4: Onset Detection system

The frequency evolution over time is obtained using the Short Time Fourier Transform (STFT), which is calculated using a Hanning window.

$$X(m,k)=\sum_{n=0}^{L-1}x(n+mH)w(n)*e^{-j(2\pi/N)k.n} \tag{4}$$

where $w(n)$ is the window that selects a L length block from the input signal $x(n)$, m and k are the frame and bin numbers respectively, and H is the hop length in samples.

Then, the frame representation in the frequency domain $X(m,k)$ is fed into a bank of FIR comb filters, which uses different delays D_i, where i is a configurable range of filter delays $[D_{min}...D_{max}]$. Next, the filter output $Y_D(m,k)$ is calculated as follows:

$$Y_{D_i}(m,k)=X(m,k)\times H(D_i,k) \tag{5}$$

where $H(D_i,k)$ denotes a FIR comb filter frequency response built with a delay D_i.

Then, the energy of each output is calculated in the frequency domain as follows:

$$E(m,D_i)=\sum_{k_i=1}^{M}\left\{\left|Y_{D_i}(m,k)^2\right|\right\} \tag{6}$$

where M denotes the FFT length.

From equation 1, it can be seen that the maximum output amplitude that the FIR comb filters can reach with g =1 is

$y_{max}(n)$ = 2*$x(n)$, which can only occur for the case of $x(n)$ = $x(n+D)$. In that case, the maximum output energy is $E(y_{max})$ = 4*$x^2(n)$. Then, by normalizing each output energy $E(m,D_i)$ with $E(y_{max})$, a measure of how similar the filter $H(D_i,k)$ is to the perfect FIR comb filter that extracts the maximum energy $E(y_{max})$ is obtained:

$$E_m(m,D_i) = \frac{E(m,D_i)}{E(y_{\max})} \qquad (7)$$

Since comb filter peaks are equally spaced along the frequency domain (see Figure 3), $E(m,D_i)$ will vary considerably depending on the spectral harmonicity of the peaks of the analyzed signal. Thus, equation 7 provides a compromise between spectral harmonicity and energy filtered, which we call "spectral fit". As an example, a FIR comb filter $H(D_i,k)$ with peaks in the magnitude response matching the harmonic peaks of a monophonic signal, will have $E_m(m,D_i)$ close to 1. In contrast, a filter with peaks that do not coincide with the bins where the energy of the signal is will have $E_m(m,D_i)$ closer to 0, which is common in the onset component of a musical signal.

Since we are interested in the deviation of $E_m(m,D_i)$ from the perfect "spectral fit", the following transformation is performed:

$$E'(m,D_i) = abs(E_m(m,D)-1) \qquad (8)$$

Thus, $E'_m(m,D_i)$ equal to 0 and 1 corresponds to the perfect and worst spectral fit respectively.

In order to obtain the onset detection function, the sum of the squared difference [17], between the maximized output energies for each delay D_i is performed for each pair of consecutive frames as follows:

$$dE(m) = \sum_{i=D_{\min}}^{D_{\max}} \left[E'(m,D_i) - E'(m-1,D_i)\right]^2 \qquad (9)$$

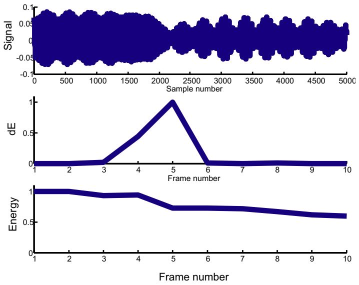

Figure 5: onset detection function of a tin whistle signal

In Figure 5, the onset detection function of a tin whistle signal (top plot) utilizing the presented approach is depicted in the middle plot. As a comparison, the energy function of the signal is also shown in the bottom plot [1]. In the FIR comb filter based approach, there is a prominent peak at the onset position; however, the energy function does not show an increase in the onset component. In order to illustrate how the onset peak arises, the $E'_m(m,D_i)$ function for the frame range m = 2 to 6 are depicted in Figure 6. The delays utilized correspond to the pitch period of the 12 notes of the third octave, and the sampling frequency is 44100 Hz. It can be appreciate that there is not a noticeable change in the functions between frames 2 and 3, and frames 5 and 6. However, there is a significant change between frames 3 and 4, and frames 4 and 5, which are the samples where the onset occurs, as can be appreciated in Figure 5.

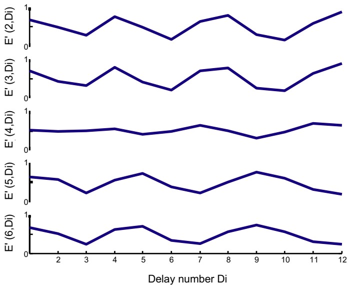

Figure 6: $E'(m,D_i)$ function for frames m = [2...6]

5. RESULTS

As mentioned in sections 1 and 2, slow onsets such as wind instruments are more difficult to detect. Figure 7 illustrates the strength of the presented approach for detecting onsets in a tin whistle signal, with its spectrogram and time domain representation depicted in the first and second row respectively. The results obtained by the use of the proposed method are depicted in the third row of Figure 7. To demonstrate the benefits of the approach, the onset detection function of four other methods is also shown. The systems utilized were the spectral difference method [17] (see the fourth row), the complex based method [15] (see the fifth row), the phase based method [14] (see the sixth row), and an energy based method [1] (see the seventh row). All functions were obtained by the use of a STFT analysis, with a frame length N and hop size H equal to 1024 and 512 samples respectively. Then, the resulting onset detection functions were normalized.

It can be seen that the onset detection function in the presented approach shows very distinct peaks at the position of the onsets. However, the other onset detection functions do not contain very clear onsets, resulting in inaccurate detections. In the analyzed signal of Figure 7, two notes are played with a slide effect, which is an inflection of the pitch (see sample ranges [8000…26000] and [48000…67000] respectively). It can be appreciated that the slide, which produces a gradual change in the amplitude and harmonicity of the frequency tracks in the spectrogram, do not alter the accuracy of the detection in the proposed approach (see third row). In contrast, the onset detection function of other methods contains several spurious peaks, which are significantly more noticeable in the second inflected note. The proposed onset detector is sensitive to sudden harmonicity changes that typically occur during the noisy onset component of the signal. However, the offset part of the signal can also have unexpected energy and harmonicity changes, which causes a spurious detection in the proposed approach at sample 67000 approximately (see row 3). The only approach that did not detect that spurious offset is the energy based system (see row 7). However, by only analyzing the sections of the signal where there is an energy increase, rapid note changes will also remain undetected (see Figure 5, bottom plot).

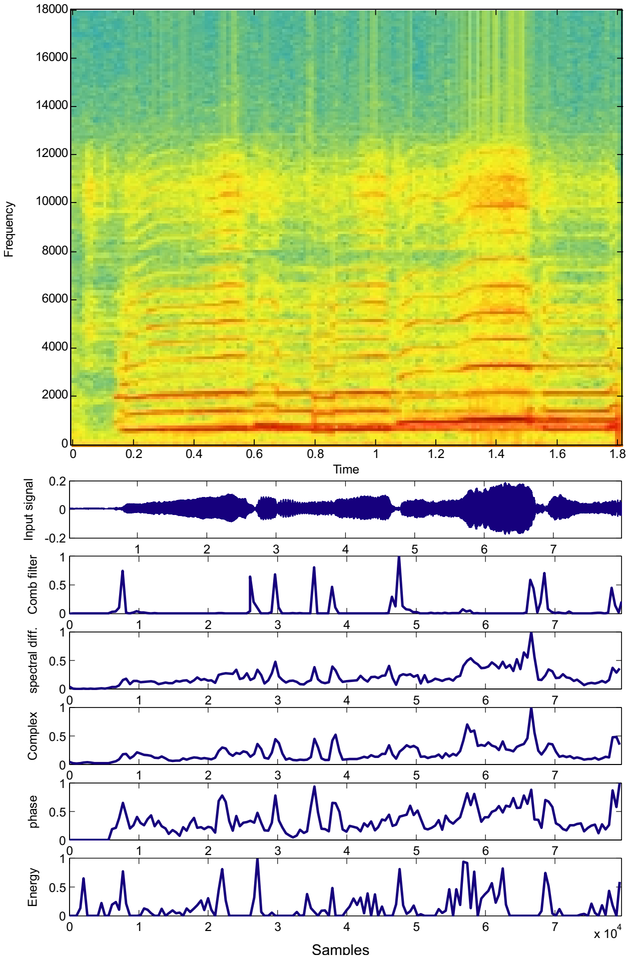

Figure 7: Onset detection methods comparison

6. CONCLUSIONS AND FURTHER WORK

A system that detects note onsets using FIR comb filters is presented. The system improves upon the performance of other onset detector approaches such as an energy based, phase based, complex based and spectral difference based onset detector, in the difficult case of detecting slow onsets. This shows that by combining the inharmonicity properties of the onset component of the signal with the energy increase utilizing FIR comb filters, the accuracy of the onset detection function is improved. To implement a multiresolution onset detector in order to obtain more accurate time precision, and investigating a method in order to avoid the detection of spurious offsets should be considered as an area for future research.

7. ACKNOWLEDGEMENTS

The authors wish to express their most sincere gratitude to Juan Pablo Bello, from Queen Mary University of London, for his kindness and generosity making the code available that made possible the comparison tests.

8. REFERENCES

[1] P. Masri and A. Bateman, "Improved Modelling of Attack Transients in Music Analysis-Resynthesis," in Proc *International Computer Music Conference (ICMC)*, 1996.

[2] E. Scheirer, "Tempo and Beat Analysis of Acoustic Musical Signals," *J. Acoust. Soc. Am.*, pp. 588-601, 1998.

[3] M. Marolt et al., "On detecting note onsets in piano music," in Proc *11th Mediterranean Electrotechnical Conference. MELECON*, 2002

[4] A. Klapuri et al., "Automatic Transcription of Music," 2001

[5] A. Klapuri and T. Virtanen, "Automatic Transcription of Musical Recordings," in Proc *Consistent & Reliable Acoustic Cues Workshop, CRAC-01*, Aalborg, 2001.

[6] M. Gainza et al., "Onset Detection and Music Transcription for the Irish Tin Whistle," in Proc *Irish Signals and Systems Conference, ISSC*, Belfast, 2004.

[7] M. Gainza et al., "Single-Note Ornaments Transcription For The Irish Tin Whistle Based On Onset Detection," in Proc *Digital Audio Effects (DAFX-04)*, Naples, 2004.

[8] D. Dorran and R. Lawlor. "Time-scale modification of music using a synchronized subband/time-domain approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Montreal, 2004.

[9] T. Virtanen and A. Klapuri, "Separation of Harmonic Sound Sources Using Sinusoidal Modeling," in Proc *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2000.*

[10] Chafe et al., "Source Separation and Note Identification in Polyphonic Music," *CCRMA Department of Music, Stanford University, California*, 1985.

[11] J. A. Bilmes, "Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Expressive Timing in Percussive Rhythm," MSc Thesis, MIT, 1993

[12] A. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," in Proc *IEEE Int. Conference on Acoustics, Speech, and Signal Processing,* 1999.

[13] B. C. J. Moore et al., "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," *J. Audio Eng. Soc.*, vol. 45, pp. 224-240, 1997.

[14] J. P. Bello and M. Sandler, "Phase-based note onset detection for music signals," in Proc. *IEEE International Conf. on Acoustics, Speech, and Signal Processing, 2003.*

[15] C. Duxbury et al., "Complex Domain Onset Detection For Musical SIgnals," in Proc *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03),* London, 2003

[16] C. Roads, *The Computer Music Tutorial*: MIT Press, 1996

[17] C. Duxbury et al., "A hybrid approach to musical note onset detection," in Proc o*f the 5th Int. Conference on Digital Audio Effects (DAFx-02)*, Hamburg, 2002