ONSET BASED AUDIO SEGMENTATION FOR THE IRISH TIN WHISTLE

Mikel Gainza[ϕ] Bob Lawlor* Eugene Coyle[ϕ]
[ϕ]Digital Media Centre, Dublin Institute of Technology, Dublin, Ireland. [mikel.gainza, eugene.coyle] @dit.ie
* National University of Ireland, Maynooth, Ireland. rlawlor@eeng.may.ie

Abstract: A technique for segmenting tin whistle audio signals according to the position of the note onsets is presented. This method focuses on the characteristics of the tin whistle within Irish traditional music, customising a time-frequency based representation for detecting the instant when a note starts and releases.
Musical ornamentation, such as cuts and strikes, are very common in Irish Traditional music and are played during the onset stage. Taking advantage of this musical feature, a novel technique for improving the onset time estimation is also presented.

1. Introduction

A musical onset is defined as the precise time when a new note is produced by an instrument, and a musical offset is the progressive release of the note.
The onset of a note is very important in instrument recognition, as the timbre of a note with a removed onset could be very difficult to recognise. Masri [1] stated that in traditional instruments, an onset is the phase during which resonances are built up, before the steady state of the signal. Other applications use separate onset detectors in their systems, like in rhythm and beat tracking systems [2], music transcriptors [3, 4, 5], time stretching [6], or music instrument separators [7, 5].

The onset detectors encounter problems in notes that fade-in, in fast passages, in ornamentations such as grace notes, trills and fast arpeggios and in glissando (fast transition between notes) or cuts and strikes in traditional music, which are discussed in section 3. Also, the physics of the instruments and recording environments can produce artefacts, resulting in a 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 focuses on the existing approaches that have dealt with the onset detection problem. In section 3 we describe the main characteristics of the Irish tin whistle and we present a method for segmenting audio based on onset detection, which takes those characteristics into consideration. Some results which validate the approach are shown in section 4 and finally, some conclusions and further work are discussed in section 5.

2. Existing Approaches

There are many different types of onsets. However, the two most common are:
A fast onset, which is a small zone of short duration of the signal with an abrupt change in the energy profile, appearing as a wide band noise burst in the spectrogram (see Figure 1). This change manifests itself particularly in the high frequencies and is typical in percussive instruments.

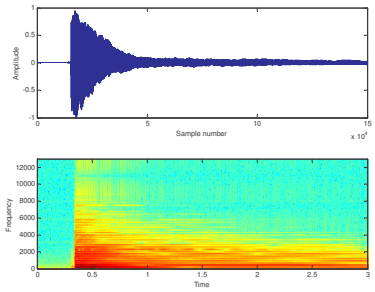

Figure 1: spectrogram of Piano playing C_4

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

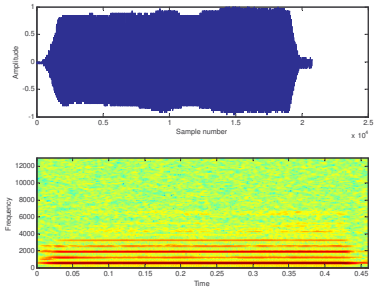

Figure 2: spectrogram of a tin whistle playing E_4

A significant amount of research on onset detection has been undertaken. However, accurate detection of slow onsets remains unsolved.
Early work which dealt with the problem took the amplitude envelope of the entire input signal for onset detection [8]. However, this approach only works for signals that have a very prominent onset, which led to the development of multi-band approaches for giving information on specific frequency regions where the onset occurs. This was first suggested by Bilmes [9], who computed the short time energy of a high frequency band using a sliding window, and by Masri in [1], who gave more weight to the high frequency content (HFC) of the signal. However, these two methods only work well for sharp onsets. Scheirer in [2], presents a system for estimating the beat and tempo of acoustic signals requiring onset detection. A 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.

The system produced good results, however, the amount of band amplitude envelopes are not enough for resolving fast transitions between notes in non percussive onsets. Klapuri [10], developed an onset detector system based on Scheirer's model. He used a bank of 21 non-overlapping filters covering the critical bands of the human auditory system, incorporating Moore's psycoacoustic loudness perception model [11] into his system. Klapuri obtained the loudness of every band peak, to combine all peaks together sorted in time, and calculated a new peak value for every onset candidate by summing the peak values within a 50 ms time window centered in the onset candidate. This approach is not appropriate for onsets that have energy in a few harmonics, because it would only produce peaks in a few bands.

Other approaches [12, 13] use phase based onset detection based on phase vocoder theory to calculate the difference between the expected and detected phase.

3. Proposed Approach

This section is subdivided into two parts: section 3.1 describes the most important aspects of the characteristics of the Irish tin whistle, and this knowledge is then used to develop an appropriate onset based segmenting system

3.1 Tin Whistle Theory

Tin whistles come in a variety of different keys. However, the most common is the small D whistle, which is used in more than 80% of Irish traditional tunes. This whistle is a "transposing instrument", which means that when it is played, the note that is heard differs from the written musical notation. For example, for the small D whistle, if a D_4 note is written on the score, a D_5 note sounds (one octave higher). To refer to a given note, this score notation will be used in this paper.

The small D key whistle is capable of playing in many different modes. Some of them require a half hole covering, which is not practical in many musical situations. Without half covering, the following modes that are very common in Irish Traditional Music can be played with the small D Whistle [14]: D Ionian (major scale) and D Mixolydian, E Dorian and E Aeolian (natural minor), G Ionian (major), A Mixolydian and A Dorian, and B Aeolian (natural minor)
If the tune is played in a key that requires half covering, like the F note in D Dorian, the player will change to a tin whistle that can play the mode without using half covering, like a C key Whistle.
Therefore, only the following notes shown in table 1 are considered in the presented algorithm:

Octave 4								Octave 5					
D	E	F#	G	A	B	C	C#	D	E	F#	G	A	B

Table 1: Full covering notes for the D tin whistle

Ornamentation plays a very important role in Irish Traditional music, giving more expression to the music altering or embellishing small pieces of a melody.
There are many different types of ornamentation in Irish traditional music: cut, strike, slide, rolls, trill, etc [14], but cuts and strikes are the ornamentation types most commonly used in Irish traditional music.
Cuts and strikes are pitch articulations: the cut is a subtle and quick lift of the finger covering its hole followed by an immediate replacement, which increases the pitch, and the strike is a rapid impact of an uncovered hole that momentarily lowers the pitch. The sound of both is very brief, and not perceived as having a discernible pitch, note or duration [14]. Therefore, they are not considered to be notes, nor graces notes, but rather are just part of the onset, providing relevant information for estimating the onset time more accurately.

3.2 System Overview

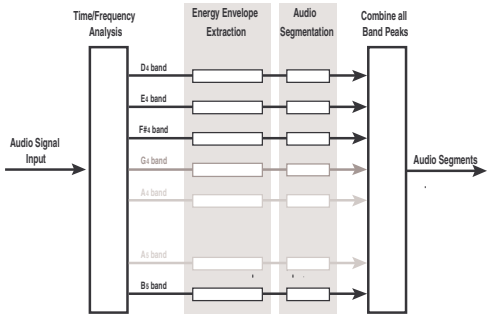

Figure 3: system overview

This section describes the different parts of the proposed onset detector system. A time - frequency analysis is first required, which splits the signal into 14 frequency bands, one band per note shown in table 1. The energy envelope is calculated for every band, which is used then to obtain the first derivative function of the envelope. To investigate the existence of onset and offset candidates, positive and negative energy changes are separated into two different functions. Then, onsets and offsets are matched together to form audio segments, and are classified in note and ornamentation segment candidates. Finally, all band segments are combined to obtain the correct onset times and note pitches.

Time-Frequency Analysis

The audio signal is first sampled at 44100 Hz. Then, the frequency evolution over time is obtained using the Short Time Fourier Transform (STFT), which is calculated using a 1024 sample Hanning window (23 ms), 50% overlap between adjacent frames and 4096 FFT length. These parameters interpolate the spectrum by a factor of 4, which is required for accuracy purposes.

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

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

Every frame is filtered using a bank of 14 band pass filters. Each band covers a logarithmic note range centered at the frequency of the notes shown in table 1.

Energy Envelope Extraction

The average energy is calculated in each band for each frame:

$$E_{av(i,n)} = \sum_{k_i=1}^{l_i} \left\{ \left| X_i(k_i, n) \right|^2 \right\} \qquad (2)$$

where X_i is the filter output of band i, k_i is i_{th} frequency bin number and l_i, is the band i length in frequency bins.

This operation smoothes the subband signal, limiting the effect of signal discontinuities. However, additional smoothing is still required, which is obtained by convolving the average energy signal with a 46 ms Half Hanning window. This operation performs a similar operation to the human auditory system, masking fast amplitude modulations but emphasizing the most recent inputs [2]. The smoothed signal after being convolved is denoted as $E_{(i,n)}$.

Audio segmentation

The first order difference of the energy envelope is calculated for each band and then, the energy increases and decreases are separated into two different vectors, $D_{E(i,n)}$ and $D_{D(i,n)}$, which are then searched for onset and offset candidate peaks respectively, that reach a predetermined threshold.
Figure 4 illustrates a G_4 note played with a cut when moving from note E_4 (top plot). Middle and bottom left plots show the $D_{E(i,n)}$ of the E_4 and G_4 bands respectively. Middle and bottom right plots show the $D_{D(i,n)}$ of the E_4 and G_4 bands respectively

Other multi-band energy based approaches [2, 10] used the same threshold for every band. However, this is not adequate for wind instruments such as the tin whistle, where strong amplitude modulations in high bands can have similar peak values as onset peaks in low bands.
Each note of a wind instrument has a different pressure range within which the note will sound satisfactory; this range increases with the frequency. Martin [15] stated that usual practice for recorder players is to use a blowing pressure proportional to the note frequency, thus the pressure increases by a factor 2 for an octave jump. We can then conclude that as with the note frequency, the general blowing pressure for different notes is spaced logarithmically. This also applies to the tin whistle, due to its acoustic similarity with the recorder.
In both cases, the threshold should also be proportional to the frequency and will have a logarithmic spacing. Then, the threshold for a band i will be:

$$T_i = T * 2^{\frac{s}{12}} \qquad (3)$$

where T is the threshold required for the band of a given note x, and s is the semitone separation between the note in the i band and the reference note x.

An onset candidate is detected if:

$$D_{E(i,n)} = E_{(i,n)} - E_{(i,n-1)} \geq T_i \qquad (4)$$

An offset candidate is detected if:

$$D_{D(i,n)} = E_{(i,n)} - E_{(i,n-1)} < -T_i \qquad (5)$$

Then, every onset candidate t_{on} is matched to the closest offset candidate in time t_{off}, where $t_{off} > t_{on}$, to form audio segments $Sg = [t_{on}, t_{off}]$
Next, according to time duration, the audio segments are split into note and ornamentation segments as follows:

$$Sg = Sg_{orn} \quad \text{if } t_{off} - t_{on} < T_e \qquad (6)$$

$$Sg = Sg_{note} \quad \text{if } t_{off} - t_{on} > T_e \qquad (7)$$

Where T_e is the longest expected ornamentation time.

As can be appreciated in figure 4, for T_e= 44 ms, T_2 =100 and T_5 = 119, a note segment will be formed in band E_4 (i =2): Sg_{note}= [$D_E(2,10)$, $D_D(2,49)$]
And an ornamentation segment will be formed in band A_4 (i =5): Sg_{orn}= [$D_E(5,7)$, $D_D(5,10)$]

Note segments in every band are combined and sorted in onset segment time order. Next, for every onset segment, only the segment that has the strongest $D_{E(i,n)}$ value between the onset and offset segment time, t_{on}, and t_{off}, is kept. The same process is repeated for the ornamentation segments.

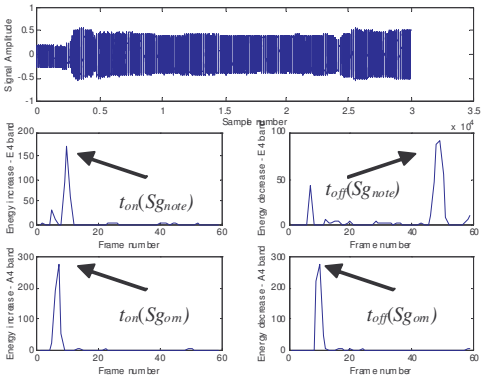

Figure 4: Cut ascending from note E_4 to note G_4 (top plot). Middle and bottom plots show the first order difference function in G_4 and A_4 frequency Band respectively: Energy increases on the left plot and the absolute value of the energy decreases on the right plot.

More accurate onset time estimation

If an ornamentation and a note segment come consecutively, $t_{off}(Sg_{orn}) \cong t_{on}(Sg_{note})$, we will consider

that the note represented in Sg_{note} was played with the ornamentation represented in Sg_{orn}. Ornamentation in Irish Traditional music is played right on the beat, providing an accurate time for when a new note starts. Then, the modified audio segment will be composed of Sg'_{note}= $[t_{on}(Sg_{orn}), t_{off}(Sg_{on})]$. The rest of the ornamentation segments, Sg_{orn}, are discarded.
As an example, the modified note segment in figure 4 will be Sg'_{note}=[$D_E(5,7), D_D(2,49)$]

The first derivative is adequate for looking for onsets. However, it is not a satisfactory technique to obtain onset time in slow onsets such as the tin whistle, which take some time to reach the peak [10].
Therefore, once the onset peak has been identified in $D_{E(i,n)}$, a more accurate onset time will be at the point before the onset peak where the function starts increasing.

4. Results

Two excerpts of Irish traditional music tunes were used for evaluating the performance of the presented system on detecting note segments and the corresponding pitches. These tunes come from Grey Larsen's book [14] with the corresponding music notation, which was very useful for verifying the results. In [16], an onset detector was presented and compared against the widely cited energy based onset detector approach of Klapuri in [10], thus consolidating the approach. The presented onset based segmentation is more challenging, since a wrong onset or offset detection in a note or ornamentation segment, would result in wrong segment detection.

The percentage of correct segment detections was calculated using the following equation [10]:

$$correct = \frac{total - undetected - spourious}{total} * 100\% \quad (8)$$

Results are shown in table 2, the first tune used (Tune 1 in table 2) is a 10 seconds excerpt of "The Cliffs of Moher [14, p337], and the second (Tune 2 in table 2) is a 11 seconds excerpt of "Willy Coleman´s gig" [14, p344]. The percentage of correct detections was very high in both tunes.

Tune	Undetected (%)	Spurious	Correct segment (%)
1	3/49 = 8.1%	1	91.9 %
2	2/46 = 4.3%	3	89.1%

Table 2: Segments detection results

5. Conclusions

A system that segments D key tin Whistle audio signals was presented. Previously, a summary of onset detector literature review was presented and the onset detector system was customised to the D key tin whistle. Also, a novel method for detecting notes played using ornamentation techniques is presented, which is used for obtaining more accurate onset estimation. The algorithm was tested for real audio signals, and the results show the strength of approach. Transcribing every type of ornamentation should be considered as an area for future research.

References

[1] Masri P. Bateman A. 1996. "Improved modelling of attack transients in music analysis resynthesis" in Proc. International Computer Music Conference-

[2] Scheirer, E., "Tempo and Beat Analysis of Acoustic Musical Signals", J. Acoust. Soc. Am. 103:1 (Jan 1998), pp. 588-601

[3] Klapuri, Virtanen. "Automatic Transcription of Musical Recordings" Consistent & Reliable Acoustic Cues Workshop, CRAC-01, Aalborg,

[4] Marolt, M. Kavcic, A. "On detecting note onsets in piano music". IEEE Electrotechnical Conference. MELECON 2002. 11th Mediterranean.

[5] Klapuri. " Multipitch estimation and sound separation by the spectral smoothness principle". IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2001.

[6] Chris Duxbury, Mark Sandler, Mike Davies "Temporal Segmentation and Pre-Analysis for Nonlinear Time-Scaling of Audio". 114th AES Convention, Amsterdam, 2003.

[7] Virtanen, Klapuri. " Separation of Harmonic Sound Sources Using Sinusoidal Modeling". IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2000.

[8] Chafe, Jaffe, Kashima, Mont-Reynaud, Smith. "Source Separation and Note Identification in Polyphonic Music". CCRMA, Department of Music, Stanford Unicersity, California, 1985.

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

[10] Klapuri A. "Sound Onset Detection by Applying Psychoacoustic Knowledge", In Proc IEEE International Conference on Acoustics, Speech and Signal Processing, 1999.

[11] Moore B., Glasberg B., Baer T. "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness". J. Audio Eng. Soc., Vol. 45, No. 4, pp. 224–240. April 1997.

[12] Bello J.P., Sandler M., "Phase-based note onset detection for music signals" ,In proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing , 2003

[13] C. Duxbury, J.P. Bello, M. Davies, and M. Sandler, "complex domain onset detection for musical signals" in Proceedings of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, September 8-11, 2003.

[14] Larsen G., "The Essential Guide to Irish Flute and Tin Whistle" Mel Bay Publications, 2003.

[15] J. Martin "The Acoustics of the Recorder". Moeck, 1994.

[16] Gainza M., Lawlor B, Coyle E." Onset Detection and Music Transcription for the Irish Tin Whistle" ISSC 2004, Belfast.