

Onset Detection and Music Transcription for the Irish Tin Whistle

Mikel Gainza^ϕ, Bob Lawlor* and Eugene Coyle^ϕ

^ϕDigital Media Centre
Dublin Institute of Technology
Dublin, IRELAND

E-mail: ^ϕ mikel.gainza@dit.ie, eugene.coyle@dit.ie

* Department of Electronic Engineering
National University of Ireland
Maynooth, IRELAND

E-mail: *rlawlor@eeng.may.ie

Abstract -- A technique for detecting tin whistle note onsets and transcribing the corresponding pitches is presented. This method focuses on the characteristics of the tin whistle within Irish traditional music, customising a time-frequency based representation for extracting the instant when a note starts and the music notation.

Results show that the presented approach improves upon the existing energy based approaches in terms of the percentage of correct detections.

I 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 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]. Also, onset detectors can be used for segmentation and analysis of acoustic signals according to the position of the onsets.

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 spurious 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 an onset detector method 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.

II 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.

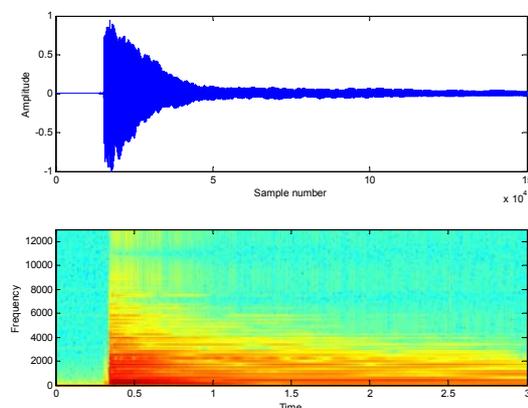


Figure 1: spectrogram of a Piano playing C₄

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.

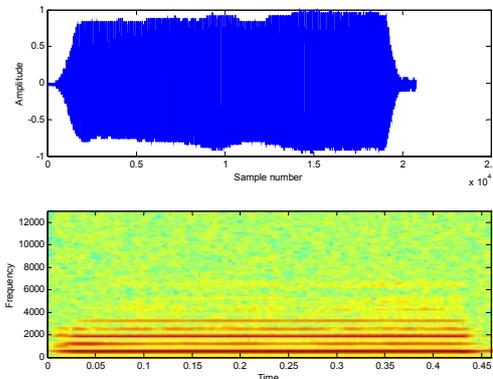


Figure 2: Spectrogram of a tin whistle playing E₄

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 psychoacoustic loudness perception model [11] into his system. To obtain the loudness of every band peak, their corresponding intensities must be first calculated. This is achieved by multiplying the peak onset value by the band center frequency, which gives more weight to high bands, thus favouring percussive onsets. Finally, the peaks in all frequency bands are combined together and sorted in time, by summing the peak values within a 50 ms time window. This approach is not

appropriate for onsets that have energy in a few harmonics, because it would only produce peaks in a few bands.

Duxbury [12] proposed a hybrid approach that uses different methods in high and low subbands for detecting different types of onsets, which can be tuned for detecting fast or slow onsets. The lowest subband (<2 kHz) used a Euclidean distance measure between successive time frames, obtaining the average energy increase over a frame.

Other approaches [13,14,15] use phase based onset detection based on phase vocoder theory to calculate the difference between the expected and detected phase.

III PROPOSED APPROACH

a) Introduction

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

b) Tin Whistle Theory

Use of the tin whistle dates from the third century A.D. [16]. However, it was not until the 1960's that the instrument started to occupy the important role in Irish traditional music that it has today.

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 D₄ note is written on the score, a D₅ note sounds (one octave higher). To refer to a given note, this score notation will be used in this paper.

The tin whistle can play in 3 octaves, but only the 2 lowest are played in Irish traditional music since the third octave sounds quite strident and shrill. Therefore, only those octaves are considered 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 [17]:

- D Ionian (major scale) and D Mixolydian
- E Dorian and E Aeolian (natural minor)
- G Ionian (major)
- A Mixolydian and A Dorian
- 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; however, it is understood in a different manner than in classical music. Ornamentation in traditional music is used for giving more expression to the music altering or embellishing small pieces of a melody. On the other hand, classical music adds music expression by adding notes to the melody.

There are many different types of ornamentation in Irish traditional music: cut, strike, slide, rolls, trill, etc [17], 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 [17]. Therefore, they are not considered to be notes, nor graces notes, but rather are just part of the onset. These articulations are selective to the player, and as stated above are not notes, therefore, they are not going to be considered as part of the music notation. However, because they are part of the onset, they provide relevant information for estimating the onset time more accurately.

c) System Overview

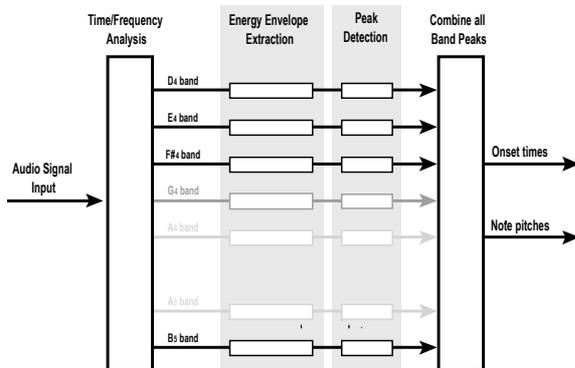


Figure 3: system overview

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. Peaks greater than a band dependent threshold in the first derivative function will be considered as onset candidates. Finally, all band peaks are combined to obtain the correct onset times and note pitches.

Time-Frequency Analysis and Multi Band Decomposition

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. The STFT is given by:

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

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

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

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

where X_i is the filter output of band i , k_i is i 's 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

$$E_{(i,n)}.$$

Peak Picking and Thresholding

Next, the first order difference of the energy envelope is calculated for each band, and peaks that reach a predetermined threshold will be considered as possible onsets. Other multi-band energy based approaches [2, 10] used the same threshold for every band. However, as can be appreciated in figure 4, this is not adequate for wind instruments such as the tin whistle. The top plot of figure 4 shows an excerpt of a tin whistle playing 3 notes: G₄, A₄ and D₄. In the middle plot it can be appreciated that setting a threshold $T < 60$ would be adequate for detecting the D₄ onset peak (middle plot). However, a slide at the ending of the G₄ note was played, producing a peak in the A₄ band (frame 26 in bottom plot) that is larger than the threshold T resulting in a false onset detection. Also, there was an amplitude modulation during the steady state region of A₄ that produced a peak (frame 49) which was close to T . Therefore, to avoid detecting spurious onsets in the A₄ band, the threshold should have a higher value.

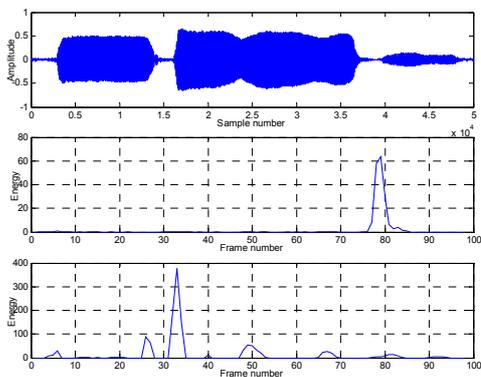


Figure 4: Excerpt of a tin whistle tune (top), D₄ frequency band (middle) and A₄ frequency band (bottom)

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 [18] 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:

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

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 will be detected if:

$$E_{(i,n)} - E_{(i,n-1)} > T_i \quad (4)$$

Combining the peak bands and ornamentation

Onset candidate peaks in every band are combined and sorted in time (frame number). If two or more peaks are located in the same, previous or next frame, we will consider that they belong to the same onset, keeping the strongest peak as final onset.

Next, a sliding 46 ms window centered at every onset candidate is applied. At this stage, two different scenarios can occur: the window contains one or two peaks.

If there is just one peak in the window, the peak frame number and the band number provide an estimation of the onset time and the new note pitch respectively (e.g.: if the peak occurs in band $i = 2$, an E₄ note is transcribed).

Two peaks within a window will mean that an articulation peak has been detected, which can occur in a different band than the one associated with the pitch, as can be appreciated in figure 5, which illustrates a G₄ note being played with a cut. The onset of the new note will be composed of:

- The articulation peak, which occurs right on the beat and gives us the onset time (see bottom plot)
- The peak in the new note band, which occurs just after the articulation peak, and which gives us the pitch of the new note (see centre plot)

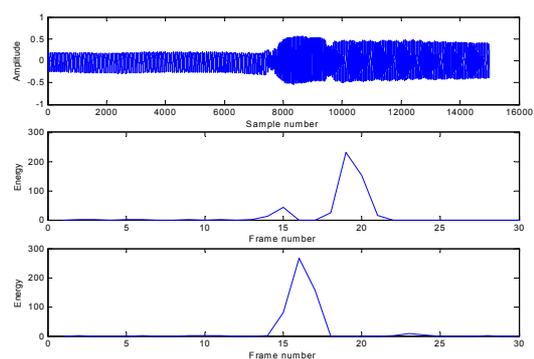


Figure 5: Cut ascending from note E₄ to note G₄ (top plot). Relative difference function in G₄ frequency Band (middle plot). Relative difference function in A₄ Band (bottom plot)

More accurate onset time estimation

The first derivative is adequate for looking for onsets, since the peaks are a good estimation of the prominence of the onset. However, is it not a

satisfactory way to obtain the onset time, especially 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 the first derivative function, the actual onset time will be at the frame before the peak where the onset stops rising.

IV RESULTS

Two excerpts of Irish traditional music tunes were used for evaluating the performance of the presented system on detecting note onset times and the corresponding pitches. These tunes come from Grey Larsen's book [17] with the corresponding music notation, which was very useful for verifying the results. To consolidate the approach, the results obtained were compared against a widely cited energy based onset detector approach, which was described by Klapuri in [10].

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

$$correct = \frac{total - undetected - spurious}{total} * 100\% \quad (5)$$

Comparison results are shown in table 2, the first tune used (Tune 1 in table 2) is a 17 seconds excerpt of "The boys of Ballisodare" [17, p134], and the second (Tune 2 in table 2) is a 16 seconds excerpt of "Bantry bay" [17, p152]. Results show that the whistle based onset detector performed better than Klapuri's system, which had some problems on detecting fast transitions between notes that occur in the same band, as in the excerpt of Tune 2 plotted in figure 6. Also, the band dependent threshold was found to be adequate for dealing with strong signal modulations. The loudness perception model did not significantly alter the onset detector system performance, since the D tin whistle note frequency range falls in a flat part of the loudness curve.

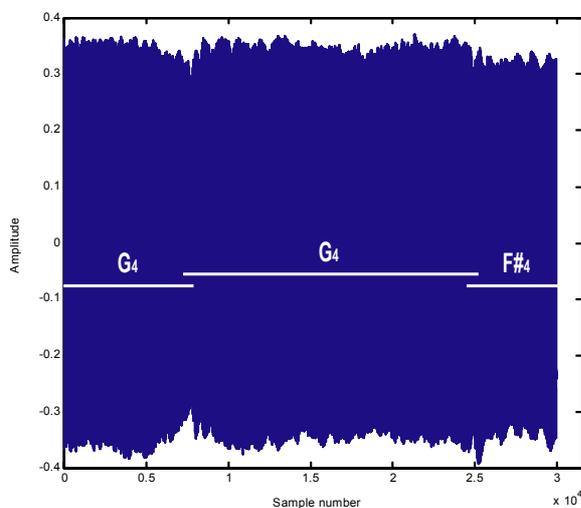


Figure 6: "Bantry bay" tune excerpt

Three spurious onsets were detected in Tune 1, which unsurprisingly occurred when a step wise descending note was played with a cut. This is the most complex cut type to play, which requires using another finger to cover a different hole, followed by lifting the cutting finger as close in time as possible. If the movement is not performed quickly enough, the peaks in the cut and pitch band are sufficiently separated to be considered as independent onsets by the system.

All detected note onsets in Tune 1 (see table 3) were transcribed into music notation correctly, however, there is one wrong pitch detection in Tune 2, which occurred when the player articulated a repeated note using a strike but without a new blowing jet, confusing the system, which identifies the strike as a new note played without articulation.

| Tune | Onset det. system | Undetected | Spurious | Correct onset (%) |
|------|-------------------|-------------|----------|-------------------|
| 1 | Tin Whistle | 0/50 = 0 % | 3 | 94 % |
| 1 | Klapuri | 4/50 = 8 % | 3 | 93.4 % |
| 2 | Tin Whistle | 0/42 = 0 % | 0 | 100 % |
| 2 | Klapuri | 2/42 = 4.8% | 2 | 85.7% |

Table 2: Onset detection comparison

| Tune | Correct pitch detections (%) |
|------|------------------------------|
| 1 | 94 % |
| 2 | 97.6 % |

Table 3: Pitch transcription results

V CONCLUSIONS AND FURTHER WORK

A system that detects note onsets and transcribes them into music notation 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 setting different band thresholds according to expected note blowing pressure was presented. The system improves upon the performance of Klapuri's onset detector, which demonstrates that customising the system according to the characteristics of the instrument, improves the onset detection accuracy.

Recently, we have become aware by personal communication of a new onset detector developed by Klapuri for analysing the meter of audio signals. We are currently studying the approach, and hope to present further comparison results in the near future.

REFERENCES

- [1] Masri P. Bateman A. 1996. "Improved modelling of attack transients in music analysis resynthesis" in Proc. International Computer Music Conference (ICMC). pp.100-103.

- [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, Denmark, September 2001.
- [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 University, 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] Duxbury, M. Sandler and M. Davies. "A hybrid approach to musical note onset detection", *in Proceedings of the 5th Int. Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany, 2002.*
- [13] 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.*
- [14] Duxbury, J.P. Bello, M. Davies, and M. Sandler, "A combined phase and amplitude based approach to onset detection for audio segmentation", *in Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-03), London, UK, 2003.*
- [15] 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.*
- [16] L.E. McCullough, " The Complete Tin Whistle Tutor". *New York Oak Publications, 1987.*
- [17] Larsen G., "The Essential Guide to Irish Flute and Tin Whistle" *Mel Bay Publications, 2003.*
- [18] J. Martin "The Acoustics of the Recorder". *Moeck, 1994.*