



Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Onset Detection, Music Transcription and Ornament Detection for the Traditional Irish Fiddle

Aileen Kelleher¹, Derry Fitzgerald², Mikel Gainza³, Eugene Coyle⁴ and Bob Lawlor⁵

^{1,3&4} Digital Media Center, Dublin Institute of Technology, Dublin, Ireland
aileen.kelleher@dit.ie, mikel.gainza@dit.ie, eugene.coyle@dit.ie

² Department of Electronic Engineering, Cork Institute of Technology, Cork, Ireland
derry.fitzgerald@cit.ie

⁵ Department of Electronic Engineering, National University of Ireland, Maynooth, Ireland
rlawlor@eeng.may.ie

ABSTRACT

By combining techniques used in previous onset detectors, a system that detects note onsets in traditional Irish fiddle tunes has been implemented. The notes detected also include the most common types of ornamentation played by the fiddle. Ornaments are notes of extremely short duration, at most a fifth the length of a regular note.

A Short Time Fourier Transform based sub-band technique, which previously gave good results for the Irish tin whistle, was modified to include a threshold approximation more suitable for the fiddle. This system has been tested on a database of real recorded fiddle tunes and good results have been achieved.

1. INTRODUCTION

An onset indicates when something new has happened in a signal. In the case of a musical signal, this could be a change in frequency, i.e. the precise time when a new note is produced by an instrument. Knowing the location of the onsets allows a signal to be segmented so

that each event can then be examined in isolation. Therefore, in order to determine what notes constitute a musical piece, an onset detector is vital. Onset detectors are used in a number of applications such as music instrument separation, time stretching and music transcription.

There are different types of onset, the most common being the fast and slow onsets. A fast onset is the easier

to detect since it has an abrupt change in the energy profile for a short duration at the beginning of a note and is very noticeable in the higher frequencies. This type is typical of the percussive instruments such as a piano or a banjo and an example of a fast onset can be seen below in Figure 1.

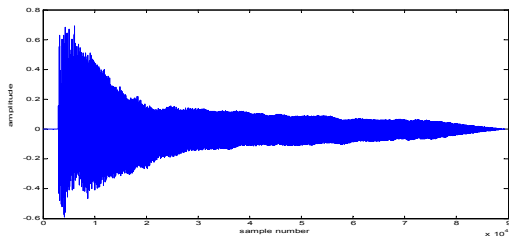


Figure 1: A single note played on a piano

The slow onset is much harder to recognize as it takes a much longer time to reach the maximum onset value and has no noticeable change in the higher frequencies. This type of onset is typical in wind instruments such as the flute and tin whistle and bowed instruments such as the fiddle. An example of a slow onset is shown below in Figure 2.

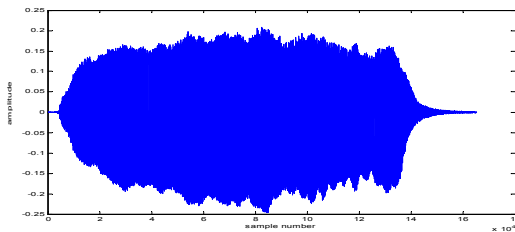


Figure 2: A single note played on a fiddle

2. EXISTING APPROACHES

A great deal of research has been carried out in the area of onset detection but it still remains an unsolved problem, particularly in the case of slow onsets. The first onset detectors considered the musical signal as a whole [1]. However this approach is only suitable for monophonic signals with very prominent onsets. Bilmes [2] introduced a sub-band approach by dividing the signal into two bands. Scheirer [3] built on these approaches by dividing the signal into six bands, one

covering each octave, which is more suitable for dealing with polyphonic signals. The amplitude envelope was extracted from each band, resulting in a series of peaks. Good results were achieved, however more bands would be required in order to resolve fast transitions between notes in non-percussive onsets.

Klapuri [4] based his onset detector on Scheirer's system. He used a bank of twenty-one filters, one for each of the critical bands of the human auditory system. The amplitude envelope for each band was calculated by convolving the signal with a half-Hanning window, this performs much the same task as the human auditory system, masking fast amplitude modulations but emphasizing most recent inputs. The first order difference function, followed by the relative first order difference function of the energy envelopes are calculated. This gives prominent peaks from which onsets can be located. Gainza [5] [6] improved upon Klapuri's sub-band technique to detect onsets in Irish tin whistle tunes. He divided the signal into fourteen bands, one for each note played by the tin whistle. He used a band dependent threshold based on tin whistle theory, which was automatically set by the system. Each note requires a certain blowing pressure in order to sound satisfactory. This pressure is proportional to the note frequency so this information was used in order to set the threshold for each band, i.e. note. Another sub-band system by Duxbury [7] proposed a hybrid approach, using different detection methods in the high and low sub-bands. The higher bands use an energy-based detector while a frequency based distance measure was used on the lower bands. He also shows that the detection threshold may be set automatically by analyzing the statistics from their detection function.

3. TRADITIONAL IRISH MUSIC

The fiddle is the most widely used instrument in Irish traditional music. It is not a native instrument as it is the same as a normal violin but the style of playing is distinctly Irish [8]. The range of notes used by traditional Irish musicians when playing the fiddle is somewhat less than that which would be used by a classical musician. In all only twenty-nine semi tones are used, those from G3 to B5. It is very unlikely that a note outside this range would be played therefore it was only these notes we were concerned with when developing our system.

Ornamentation is extremely important to the character and style of traditional Irish music. It is a feature shared with classical music, used by both to embellish a note or passage of music. This is where the similarities between the two genres end. In classical music an ornament is considered a note or group of notes in their own right whereas in traditional Irish music it is not considered to be a note, merely an alteration or embellishment belonged to a parent note. These alterations and embellishments are created mainly through the use of special articulations and inflections, not through the addition of extra, ornamental or grace notes [9].

There are many different types of ornaments used in traditional Irish music. Examples of single note ornaments include the cut, strike and slide while the double cut, roll and crann are multi-note ornaments. The cut and double cut are the most common types for this particular instrument and are the two ornaments our system was required to identify.

4. PROPOSED SYSTEM

The proposed system uses a sub-band approach to detect the onset of each note and ornament. Twenty-nine bands are used; one for each of the semi tones in the range of notes used by the fiddle.

4.1. Frequency Analysis

Firstly the signal was sampled at 44100Hz. Then a time frequency representation of the signal was obtained using the Short Time Fourier Transform (STFT), which is given by:

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

where $x(m)$ is the signal, n the frame number, H the hop length and $w(m)$ the window of length L . A 1024 sample Hanning window with 50% overlap between frames and a 4096-point Fast Fourier Transform (FFT) were used. Each frame of this frequency representation was filtered using a bank of 29 logarithmically spaced band pass filters, one for each note in the fiddle range.

To normalize the signal, the average energy of each band was calculated:

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

Additional smoothing was also required; this was achieved by convolving the energy envelope of each band with a half-Hanning window. This operation performs in much the same way as the human auditory system, masking fast amplitude modulations and emphasizing most recent inputs [4].

The first order difference function of the resulting signal from each band was then calculated to give a series of peaks from which possible onsets of notes and ornaments could be located.

4.2. Automatic Threshold

In order to determine whether a peak is a possible onset candidate, a threshold was chosen for each band. This was done automatically by analyzing the statistics of the energy envelope of each band [7].

There is a certain point in each band where a peak equal to or greater than this value is more than likely an onset. It is possible to find this point using the probability distribution function (pdf) of the energy envelope of each band. The second order difference function of this pdf was calculated giving a maximum as shown in Figure 3 below.

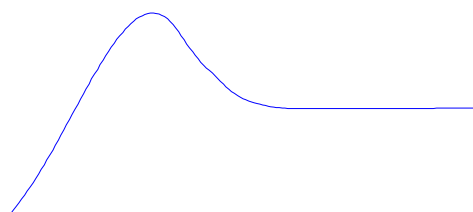


Figure 3: The second derivative of the energy envelope pdf

This maximum was chosen as the band threshold. Every value above this threshold in the first order difference function of the energy envelope of the relevant band should be an onset peak.

4.3. Band Combining

Each band is compared with the other 28 to determine possible onsets. A peak in the current frame of the current band is compared with the previous, current and next frame of all other bands. If there are no peaks in the other bands, the current frame is possibly an onset. Otherwise the peak with the greatest magnitude is the possible onset. All bands are then combined to give a series of possible onset candidates in time order. The onset locations are adjusted to be three frames before those detected. We have observed in practice that this operation gives a similar result to obtaining the relative difference function of the signal and gives a more accurate onset time, as there is a gradual slope to each peak maximum.

A second approximation of the onset locations is obtained in a simple summation of the energies in each frame of an STFT of the original signal. This is intended to pick up any onsets that may have been missed. Finally, both sets of onsets are combined with multiples at the same location eliminated.

4.4. Note and Ornament Identification

To determine the fundamental frequency of each onset, the original wave file is analyzed. This is done by performing a 4096-point Fast Fourier Transform on a 4096 sample section after one third of each note has been played. In cases where the note is of a very short duration, i.e. a possible ornament, the analysis occurs 3 samples after the note begins. Delaying the analysis gives a more accurate result in the case of slow onsets, allowing each note to reach a steady state.

Two note onsets, x and y , are expected to be a certain minimum distance, D , apart, i.e.:

$$y - x \geq D \quad (3)$$

If this is not the case, there are two possible reasons. If both onsets have the same frequency, y is a spurious onset and x is the onset time of a note. When both onsets are of different frequencies, x is the onset of a

single note ornament, i.e. a cut and y is the onset of its parent note. Another possible scenario is that a third onset, x_i , is located between x and y . In this case, if equation (3) still applies and x_i has a different frequency to x and y , then x and x_i are the onsets of a double cut and y the onset of its parent note. If x , x_i and y all have the same frequency then x is taken as a note onset, while x_i and y are taken as spurious onsets.

The result of this process is a series of note and ornament onsets of varying frequencies, which should correspond with those contained in the original wave file.

5. RESULTS

Real live recordings from Matt Cranitch's Irish Fiddle Book [8] were used to test the system. Ten excerpts from ten different tunes were selected. All but three of these excerpts contained single or double note ornaments and were, on average, ten seconds long. As these were real recordings played on a fiddle, it was found that what was played did not always exactly match the notation. This was picked up by the detection system and was only noticeable to the human ear upon very close inspection of the waveform. On that basis it was thought it would be more accurate to compare the detection results to what was actually played rather than the given notation.

The results are presented in the following two tables, the first indicating detection results for regular notes and the second for ornamentation. In Table 2, c and dc represent a cut and double cut respectively.

Tune	Undetected	Wrong Frequency	% Correct
1	0	0	100
2	2/35	0	94
3	1/34	0	97
4	3/26	0	88
5	4/38	0	89

6	3/45	0	93
7	2/34	1/34	91
8	0	0	100
9	3/40	0	93
10	0	3/36	92

Table 1: Onset detection results

Tune	Ornaments	Detected	% Correct
1	1 c 1 dc	1 c 1dc	100
2	2 c	2c	100
3	6 dc	3c	25
4	1 c	1 c	100
5	1dc	1c	50
6	4c	4c	100
7	1 c	0	0

Table 2: Ornamentation Results

The system detected regular onsets quite well with an average accuracy of 94% in the tunes that were tested. Notes detected at the wrong frequency were a semi-tone away from what they should have been. The cut, which is a single note ornament, posed no major problem to the system. However, it was found that the double cut was more difficult to detect. This is a double note ornament and the system often failed to detect the onset of the second note, as it was generally less than 10ms long. The first note of a double cut is usually of the same frequency as its parent note, which meant that it would be treated as a spurious onset in the selection process if the onset of the second note of the double cut

was not detected. In other cases, a double cut was detected as a cut. In general the average accuracy of the ornament detection was 68%.

6. CONCLUSIONS

A system for detecting the onsets and frequencies of notes and ornaments in traditional Irish fiddle tunes was implemented and tested. A sub-band technique was combined with a statistical threshold approximation and good results were achieved in the case of regular notes and single note ornaments.

Future work will concentrate on trying to improve the detection of multi-note ornaments, such as the double cut, possibly through the use of improved temporal resolution near the onsets. It is hoped that the system will perform equally well on other traditional instruments, such as the flute and tin whistle, and future work will include testing the system on these instruments.

7. REFERENCES

- [1] Chafe, "Source separation and note identification in polyphonic music," CCRMA, Department of Music, Stanford University, 1985.
- [2] Bilmes, "Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning and Reproducing Expressive Timing in Percussive Rhythm," MSc Thesis, MIT, 1993.
- [3] Scheirer, "Tempo and beat analysis of acoustic musical signals," Journal of the Acoustical Society of America, 103:1 (Jan 1998) pp. 588-601.
- [4] Klapuri. "Sound Onset Detection by Applying Psychoacoustic Knowledge," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1999.
- [5] Gainza, "Onset Detection and Music Transcription for the Irish Tin Whistle," in Proceedings of the Irish Systems and Signals Conference, Belfast, 2004.

- [6] Gainza, "*Single Note Ornamentation Transcription for the Irish Tin Whistle Based on Onset Detection*," in Proceedings of the Digital Audio Effects Conference, Naples, 2004.
- [7] Duxbury, "*A Hybrid Approach to Musical Note Onset Detection*," in Proceedings of the Digital Audio Effects Conference, Hamburg, 2002.
- [8] Cranitch, "*The Irish Fiddle Book*," Ossian, 2001.
- [9] Larsen, "*The Essential Guide to Irish Flute and Tin Whistle*," Mel Bay Publications, 2003.