AUTOMATING ORNAMENTATION TRANSCRIPTION

Mikel Gainza, Eugene Coyle
Audio Research Group (Dublin Institute of Technology)
[mikel.gainza,eugene.coyle]@dit.ie

ABSTRACT

A novel technique for detecting single and multi-note ornaments is presented. The system detects audio segments by utilising an onset detector based on comb filters (ODCF), which is capable of detecting very close events. In addition, a novel method to remove spurious onsets due to offset events is introduced. The system utilises musical ornamentation theory to decide whether a sequence of audio segments correspond to an ornamentation musical structure. In order to evaluate the results, a database of signals produced by different players using the three different instruments has been utilised. The results represent a step forward towards fully automating ornamentation transcription.

Index Terms— Music, acoustic signal analysis

1. INTRODUCTION

Ornamentation techniques are utilised for giving more expression to the music by altering or embellishing small pieces of a melody. There is no general agreement in the use of specific symbols to transcribe ornamentation, where its notation and understanding has considerably varied across centuries [1]. There are many different types of ornaments; where grace notes, appoggiaturas, mordant, trills, turns and rolls are only a few examples.

Numerous approaches perform pitch detection, including models based on detecting the periodicity of the time or frequency domain [2], auditory modelling [3], knowledge modelling [4] or data representations [5]. However, the detection of ornamentation remains an open-field of research. In [6], a single-note ornamentation detection system customised to the characteristics of the tin whistle was presented. The system is limited to the detection of just cuts and strikes. A method that uses independent component analysis (ICA) to detect piano trills is presented in [7]. A more general approach to detect different types of ornamentation is presented here. The model utilises a very accurate onset detector [8], which detects very close events that typically occur when ornamentation is utilised. By defining a set of rules that describe different types of ornamentation, the transcription of these ornaments is performed. The presented ornamentation transcription system is applied to the context of Irish traditional music, in which ornamentation plays a very important role. This represents a very challenging context, since the sound duration of the ornaments in Irish traditional music is very brief. In this case, ornaments modify the note they embellish, and only one note will be finally heard [9]. However, by defining a different set of ornaments, the approach could be applied to any type of music.

Section 2 describes the different blocks that comprise the ornamentation detector. In Section 3, the ornamentation detector is applied to the context of Irish traditional music by transcribing its most common ornaments. Next, a set of results that evaluate the ornamentation detector followed by a discussion of the obtained results is provided is Section 4. Finally, conclusions regarding the ornamentation transcription system are presented in Section 5.

2. SYSTEM DESCRIPTION

The different parts of the ornamentation transcription system presented here are depicted in Figure 1. Firstly, the onset detection block is described, from which a vector of onset candidates is obtained. Next, spurious onset detections due to offset events are removed. Following this, audio segments are formed and divided into note and ornamentation candidate segments. Next, the pitch of the audio segments is estimated. Finally, single and multi-note ornaments are transcribed in Section 4.

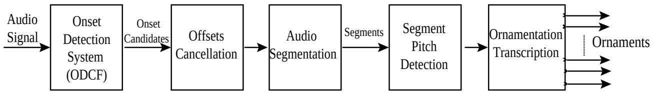

Figure 1: Ornamentation transcription system

2.1. Onset detection based on comb filters (ODCF)

Existing onset detectors utilise energy [10, 11], phase [12] or combined phase and energy signal properties [13, 14] to generate an onset detection function (ODF). In contrast, the onset detection system presented in [8], denoted as ODCF, utilises the harmonicity changes of the signal by using a bank of FIR comb filters, which also have a harmonic type of magnitude response. The filterbank utilises 12 comb filters, which delays cover the 12 semitones of octave 4. FIR comb filters can be efficiently implemented in the time domain. Thus, the harmonicity measure of [8] applied to a signal $x(n)$ by using a comb filter with delay D will be given by:

$$E(m,D) = \frac{\sum_{n=D+1}^{N}\left[x(n)+x(n-D)\right]^2}{4*\sum_{n=D+1}^{N} x^2(n)} \qquad (1)$$

where m and N are the frame number and its length respectively. Next, as in [8], the squared difference between the harmonicity measures for each delay D_i of the filter bank is performed for each pair of consecutive frames, which generates an ODF:

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

The method relates the harmonicity detection to the energy of the frame being analysed. This is suitable for detecting slow onsets, and provides a more accurate onset estimation time than other approaches. The approach is robust for dealing with amplitude modulations, e.g. if the energy of the signal changes between successive frames (but not its harmonicity) the onset detection function (ODF) remains stable. In addition, the method is robust to frequency modulations that gradually occur in the signal, since the signal harmonicity does not change considerably between frames. Considering Figure 2, where a signal excerpt containing a roll played by a flute is depicted in the top plot. The ODF of the signal generated by utilising the ODCF is depicted in the bottom plot. It can be seen that the ODCF provides a distinctive peak at the location of the new events in the signal.

Since the harmonicity measure is calculated based on the energy content of the frame, small frame lengths such as $N = 1024$ or long hop sizes such as $L = 1024$ samples, can be oversensitive to signal changes. Thus, the pair $L/N = 512/2048$ samples provides a good compromise between robustness and sensitivity to signal changes.

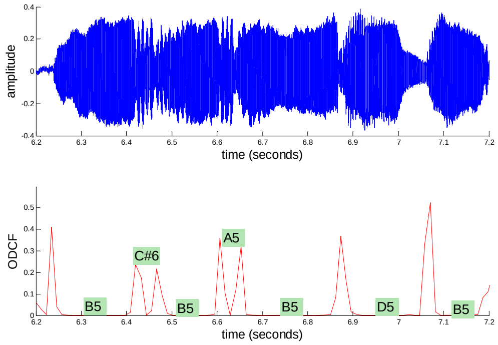

Figure 2: B5 roll - D5 - B5 sequence played by a flute

Generally, ODFs are smoothed by using a low-pass filter in order to avoid multiple spurious onset detections [10]. However, in the case of rapid ornaments, the use of smoothing will merge the successive ornamentation and note onsets into one unique peak. Consequently, no smoothing operation is performed in the presented ornamentation detector.

Peaks in the ODF that reach a given threshold are selected as onset candidates. In order to set the threshold, a method based on the standard deviation is presented. This measure provides an estimation of how the values of a signal deviate from the mean of the signal [15]. When an onset occurs, its value in the ODF is significantly prominent. Thus, the onset peak value deviates from the mean of the ODF more than its standard deviation. By using a sliding rectangular window length L centred at each frame number m of the ODF, the threshold will dynamically vary as follows:

$$T(m) = \overline{W} + \left(\frac{1}{L} \sum_{i=-L/2}^{L/2+1} (Wi - \overline{W})^2 \right)^{1/2} \quad \text{where } \overline{W} = \frac{1}{L} \sum_{i=-L/2}^{L/2+1} Wi \qquad (3)$$

where Wi are the values of ODF within the rectangular window

The use of the sliding window provides a dynamic threshold, which value varies according to the statistical content within the window. However, in very slow tunes such as “slow airs” in traditional Irish music, the separation between consecutive onsets can be long. In this case, the threshold will provide low values which produce spurious detections in that region. This limitation is overcome by setting a static and a dynamic component in the threshold, which values correspond to the mean of the entire ODF and to the standard deviation of the window respectively:

$$T(m) = \frac{1}{M} \sum_{j=1}^{M} ODF_j + \left(\frac{1}{L} \sum_{i=-L/2}^{L/2+1} (Wi - \overline{W})^2 \right)^{1/2} \quad \text{where } \overline{W} = \frac{1}{L} \sum_{i=-L/2}^{L/2+1} Wi \qquad (4)$$

where M is the length of the ODF

This threshold structure resembles to the method utilised in [12, 16], which is based on dynamic median measures of the ODCF. However, in [12, 16] the static component δ *is* set by obtaining onset detection results for different δ values. Then, the results are compared against a database of hand labelled onsets. Finally, the δ that obtains the best detection results is chosen. By contrast, Equation 2’s thresholding method is fully automated.

Onset detectors [10, 11] centred a sliding window length 50ms at each onset candidate. The most prominent candidate is maintained, while the remaining onset candidates are assumed to belong to the same onset and so are discarded. In the presented ornamentation detector, both ornamentation and note events can be separated by less than 40 ms, and using such window will cancel one of the candidates. Consequently, no window is utilised to combine onset candidates in this approach.

2.2. Offsets cancellation

The offset part of a signal also contains unexpected harmonicity changes, which can cause spurious onset detections. This can be seen in Figure 3, where the middle plot depicts the ODF generated by the ODCF of the tin whistle signal depicted in the top plot.

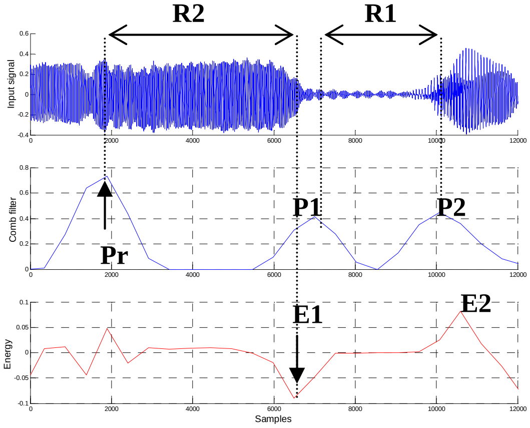

Figure 3: slow offset-onset transition

This problem is solved by applying the following method; firstly, the existence of an energy decrease peak E1 successively followed by an energy increase peak E2 is first investigated. This is the case in Figure 3, where the bottom plot shows the energy variation of the top plot signal. Next, if two peaks P1 and P2 arise in the ODF within the above mentioned transition, the first peak P1 is considered to be an offset candidate.

However, this scenario can also arise in the case of three notes connected in legato, where the second note in time order has lower energy content than the other two notes. In this case, a correct onset will be estimated as a spurious onset.

In order to differentiate between the two scenarios, a measure of the noise content of the audio signal is calculated by using a method described in [17]. First, the energy E1 and E2 of the frequency range [1:3000] Hz and [18000:21000] Hz is respectively obtained. If a note has been played, it is expected that E1 has a much higher value than E2. Thus, by performing No = E1/E2, a measure of the noise content of the signal is obtained.

The method is applied to the following two regions within the audio signal:

1. R1: region comprised in between P1 and P2 peaks in the ODF (see Figure 3), which provides a *No1* value.
2. R2: region comprised by E1 and the previous onset peak in the ODF (*Pr* in Figure 3), which provides the *No2* value.

Finally, the noise measure of both *No1* and *No2* regions are compared in order to investigate if the value of *No1* is significantly greater than *No2*:

$$No2 > Td * No1 \qquad (5)$$

where *Td* is set to a high value If the condition is fulfilled, the offset represented by P1 is removed as an onset candidate.

2.3. Audio segmentation

As in [18], every onset candidate on_n is matched to the next onset candidate in time order on_{n+1} to form audio segments Sg_n = [on_n, on_{n+1}]. Next, a table of audio segments is formed, where the second and third columns denote the beginning and ending of the audio segments. As an example, Table 1 shows the audio segments of the signal depicted in Figure 2.

n	*on*$_n$ *(sec)*	*on*$_{n+1}$ *(sec)*	*Sgn*	*P(n)*	*SNOr*	*MN Or*
1	6.235	6.42	*note*	*B5*		*Roll*
2	6.42	6.467	*orn*	*C#6*	*cut*	*Roll*
3	6.467	6.606	*note*	*B5*	*cut*	*Roll*
4	6.606	6.653	*orn*	*A5*	*str*	*Roll*
5	6.653	6.873	*note*	*B5*	*str*	*Roll*
6	6.873	7.07	*note*	*D5*		
7	7.07	…	*note*	*B5*		

Table 1: Table of audio segments of Figure 2 (top plot)

Next, according to time duration, the audio segments are split into note and ornamentation segment candidates as follows:

$$\begin{aligned} Sg_n &= orn & \text{if } on_{n+1} - on_n &< Te \\ Sg_n &= note & \text{if } on_{n+1} - on_n &> Te \end{aligned} \qquad (6)$$

where *Te* is the longest expected ornamentation time for an experienced player, which has been analytically set to *Te= 70ms.* The Sg_n segment type *is* shown in the fourth column of the audio segments table, as it can be seen in Table 1.

2.4. Segment pitch detection

In order to obtain the pitch of the audio segments, a similar method to [Brown 92]'s is utilised. Following this, the fundamental frequency estimation is refined by using parabolic interpolation [18]. The pitch of each audio segment Sg_n is shown in the fifth column of Table 1, and is denoted as *P*(*n*).

3. ORNAMENTATION TRANSCRIPTION

The system detects single-note ornaments (SNO) by utilising musical ornamentation theory [9] to establish a set of rules to decide whether a note has been played with SNO. Finally, multi-note ornaments (MNO) are formed by combining the estimated SNO and pitch information.

3.1. Single-note ornaments transcription (cuts and strikes)

- **The cut** momentarily increases the pitch. By considering Figure 2 example, it can be seen that the second and third segments in Table 1 are an ornamentation and a note segment. In addition, *P*(2)= *C#6* is higher than *P*(3) = *B5*. Consequently, B5 has been ornamented with a cut in *C#6,* and both segments together form a cut segment.
- **The strike** separates two notes of the same pitch by momentarily lowering the pitch of the second note. A strike ornament that separates two notes is also present in Figure 2 example. From Table 1, it can be derived that the fifth segment is a B5 note, which is separated from another B5 note by using the strike represented by the fourth segment.

3.2. Multi-note ornamentation transcription

Cranns and rolls are formed by combining ornamented and unornamented slurred notes of the same pitch:

- **The roll** is formed by a note followed by a cut segment, and a strike segment. By considering Table 1, it can be seen that the combination of a B5, a cut segment and a strike segment form a roll, where the three note segments have the same pitch B5. The **short roll** version removes the first unornamented note.
- **The crann** segment structure is similar to the roll. The difference lies in the use of only cuts to ornament the notes. The **short crann** removes the first unornamented note
- **The shake** is a four notes ornament formed by rapid alterations between the principal note and a note a whole or a half step above it [9]. It commences with the three ornaments and finishes with the principal note. An example of a shake can be seen Figure 4 (top plot), where an excerpt of a tin whistle tune is depicted. In the bottom plot, the ODF generated by the ODCF is also depicted. By obtaining the pitch of those segments, a sequence of three ornaments (F#5, E5, F#5) and the principal note again E5 is obtained, which corresponds to a shake ornament.

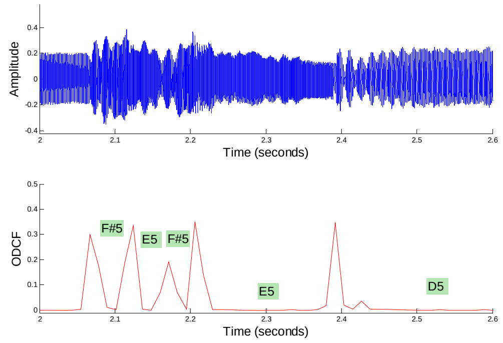

Figure 4: example of a shake played by a tin whistle

4. RESULTS AND DISCUSSION

In order to analyse the performance of the ornamentation transcriptor, two different tests have been performed. The tunes were selected from CD recordings as well as from informal live session recordings. First, a set of tin whistle excerpts is utilised in the evaluation. Next, the robustness of the algorithm in detecting ornaments produced by other instruments, including the flute and the pipe, is also investigated. In both cases, both SNO and MNO are transcribed.

The accuracy of the detection is obtained by calculating two different measures *pGP* and *acc.* The value of *pGP* (percentage of good positives) is obtained by dividing the number of correct detections by the total of ornaments in the database. A correct single-note ornamentation detection should be separated by less than 50ms from the commencement of the ornamentation, which has been previously hand-labelled. In the multi-note ornamentation case, a more flexible distance of 200ms is utilised, since two different multi-note ornaments will rarely fall within that distance. The value of *acc* is obtained by using the following equation [10]:

$$acc = \frac{total - FN - FP}{total} \qquad (7)$$

where total, *FN* and *FP* are the total of ornaments, false negatives (undetected) and false positives (spurious). It should be noted that in both cases, the ornament database is smaller than the actual notes played, which increases the difficulty of the testing method.

4.1. Test 1: tin whistle signals

The tin whistle database comprises 11 excerpts of Irish traditional music played by three different players. 493 notes comprise the database, from which 86 were single-ornamented. Also, 22 multi-note ornaments were played. The results are shown in Table 2.

	pGP **(%)**	***FN***	***FP***	***acc*** **(%)**
Single-Note	63/86 = 73.26	23	11	60.47
Multi- Note	9/22 = 40.91	13	0	40.91

Table 2: Orn. detection results for the tin whistle database

In Table 2, it can be seen that both measures *pGP* and *acc* provide high percentage results in the single-note ornamentation case. The majority of *FP* errors are due to incorrect cut detections before the real onset in the informal live recordings excerpts. These recordings have a more noisy nature than the selected CD recordings. In addition, a large number of *FN* errors is detected, which is due to repeating strikes. When playing this type of ornament on a wind instrument, the only movement of the players fingers is to rapidly cover the first uncovered hole without interrupting the flow of air [9]. Due to the extreme brevity of the strike and the high responsiveness of the tin whistle, strikes can be missed (*FN* error) by the ornamentation detector. Since the strike separates the second and third notes of the roll, this error also affects the multi-note ornamentation results.

4.2. Test 2: Flute and Pipe signals

A signal database comprised by 7 excerpts of flute and pipe excerpts coming from 4 different players is utilised. The database consists in 290 notes, from which 36 were single-ornamented. 5 MNO were also played. The results are shown in Table 3.

	pGP **(%)**	***FN***	***FP***	***acc*** **(%)**
Single-Note	31/36 = 86.11	5	10	58.33
Multi- Note	4/5 =80	1	0	80

Table 3 Orn. detection results for the flute and pipe database

As it can be seen in Table 3, the percentage of *pGP* in the SNO case is also high. As in the tin whistle test, the problem of detecting incorrect cuts in noisy parts before the onset is also manifested in this evaluation. This has the effect of increasing *FP* and consequently degrading the percentage of *acc*. The *pGP* and *acc* measures for the multi-note ornamentation case is also high. (80%) In the multi- ornamentation evaluations, both Test1 and Test2 provide a *FP* = 0. This is explained by the large number of conditions required to estimate a correct multi-note detection.

5. CONCLUSIONS

A novel ornamentation detector has been presented, which focuses on the accuracy of the onset detection results for slow onsets provided by [8]. The system also incorporates a novel thresholding method and a novel offset cancellation method. The ornamentation detector has been applied to highly ornamented Irish traditional music. However, by creating new rules to transcribe the detected audio segments, the approach could be configured to detect ornaments in other types of music. Thus, ornaments such as grace notes, mordent, trills, appoggiaturas and turns can also be detected. The system has been evaluated by using two different databases of different instruments, players and recording types. Considering the difficulty of the task, the results are relatively high in both single and MNO. However, some limitations of the approach have been identified in Section 5, which warrants future research.

6. ACKNOWLEDGEMENTS

Work supported by European Community under the Information Society Technologies (IST) programme of the 6th FP for RTD - project EASAIER contract IST-033902

7. REFERENCES

[1] Kreitner, K., Ornaments. 2006. http://www.grovemusic.com,

[2] Brown, J.C., Musical fundamental frequency tracking using a pattern recognition method. JASA 1992. 92(3): p. 1394-1402.

[3] Klapuri, A., Signal Processing Methods for the Automatic Transcription of Music. 2004.

[4] Kashino, K., et al. Organization of hierarchical perceptual sounds: In Computational auditory scene analysis workshop, intl joint conferences on artificial intelligence. 1995.

[5] Smaragdis, P. and J.C. Brown. Non-negative matrix factor for polyphonic music transcription. In IEEE WASPAA. 2003.

[6] Gainza, M., B. Lawlor et al..Single-Note Ornaments Transcription For The Irish Tin Whistle Based On Onset Detection. In DAFX-04. Naples, Italy.

[7] Brown, J. and P. Smaragdis, Independent component analysis for automatic note extraction from musical trills. JASA 2004.

[8] Gainza, M., B. Lawlor, and E. Coyle. Onset Detection Using Comb Filters. In IEEE WASPAA. 2005.

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

[10] Klapuri, A. Sound onset detection by applying psychoacoustic knowledge. in IEEE ICASSP. 1999.

[11] Masri, P. and A. Bateman. Improved Modelling of Attack Transients in Music Analysis-Resynthesis. In ICMC 1996.

[12] Bello, J.P. and M. Sandler. Phase-based note onset detection for music signals. In IEEE ICASSP. 2003.

[13] Duxbury, C., et al. Complex Domain Onset Detection For Musical SIgnals. In 6th Int. Conference on DAFx-03. London

[14] Duxbury, C., M. Davies, and M. Sandler. A hybrid approach to musical note onset detection. In 5th DAFx-02. Hamburg,

[15] Pal, N. and S. Sarkar, Statistics: Concepts and Applications. 2005: Prentice Hall of India.

[16] Bello, J.P., et al., On the use of phase and energy for musical onset detection in the complex domain. Signal Processing Letters, IEEE, 2004. 11(6): p. 553 - 556.

[17] Zolzer, U., Chapter 9 - Spectral Processing, in DAFX - Digital Audio Effects. 2002, John Wiley & Sons