AUTOMATIC TRANSCRIPTION OF ORNAMENTED IRISH TRADITIONAL FLUTE MUSIC USING HIDDEN MARKOV MODELS

Peter Jančovič[1] **Münevver Köküer**[2,1] **Wrena Baptiste**[1,2]
[1] School of Electronic, Electrical & Systems Engineering, University of Birmingham, UK
[2] School of Digital Media Technology, Birmingham City University, UK
p.jancovic@bham.ac.uk, m.kokuer@bham.ac.uk

ABSTRACT

This paper presents an automatic system for note transcription of Irish traditional flute music containing ornamentation. This is a challenging problem due to the soft nature of onsets and short durations of ornaments. The proposed automatic transcription system is based on hidden Markov models, with separate models being built for notes and for single-note ornaments. Mel-frequency cepstral coefficients are employed to represent the acoustic signal. Different setups of parameters in feature extraction and acoustic modelling are explored. Experimental evaluations are performed on monophonic flute recordings from Grey Larsen's CD. The performance of the system is evaluated in terms of the transcription of notes as well as detection of onsets. It is demonstrated that the proposed system can achieve a very good note transcription and onset detection performance. Over 28% relative improvement in terms of the F-measure is achieved for onset detection in comparison to conventional onset detection methods based on signal energy and fundamental frequency.

1. INTRODUCTION

Automatic transcription of music is concerned with converting an acoustic signal into a symbolic representation that provides the information on individual notes played and possibly also other higher-level information about the structure of music. Over the last decade, there has been a considerable research interest in this field. Although most of the current research is devoted to polyphonic music transcription, transcription of monophonic music is still of interest due to existing large amount of real-world monophonic music of specific properties. This paper deals with the transcription of notes and detection of their onsets in monophonic flute recordings of Irish traditional music that contains ornamentation. Ornamentation is used extensively in Irish traditional music by players of all melody instruments. Ornaments are notes of a very short duration. They are central to the style of the performer, adding to the liveliness and expression of the music.

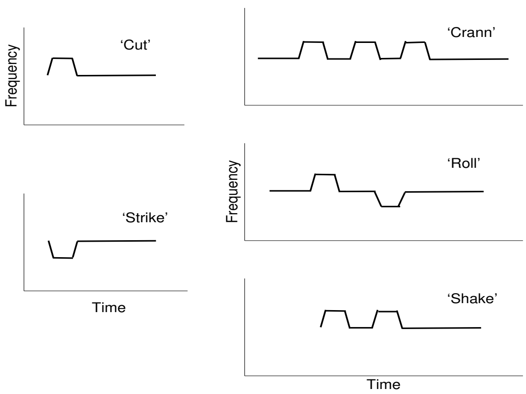
A wide range of different approaches for automatic music transcription have been proposed. A variety of algorithms for estimating the fundamental frequency (F_0), e.g., [4, 10, 16], were employed in transcription of music, e.g., [1, 5, 10]. As the F_0 estimation may suffer from making octave errors, music transcription systems typically employ some way of temporal filtering or post-processing. The use of hidden Markov models (HMMs) for post-processing was presented in several works. In [2, 15] the sequence of pitch salience and onset strength or energy difference of adjacent signal frames were used as the input features to HMMs. In [6], the acoustic signal was first segmented by applying an onset detection algorithm and then HMM was used to track note candidates. Bayesian modeling that exploits knowledge of musical sound generation was proposed in [3, 9] and applied for piano transcription. Recently, several methods based on learning of a model / classifier of notes were presented, e.g., [7, 14]. In [14], a support vector machine classifier, trained on spectral features, was used to clasify frame-level note instances and the classifier output was then temporally smoothed using a note level HMM to perform transcription of piano recordings. Modelling of a time-frequency representation of audio as a sum of basic elements representing the spectrum of a single note was presented in [7]. The transcription of ornamented Irish traditional flute music was investigated at the level of onset and ornament detection in [8, 11]. In both works, the presented ornament detection system was based on detecting onsets and using rules of musical ornamentation. An energy-based onset detection algorithm was employed in [8], while a comparison of two energy- and F_0-based onset detection algorithms was performed in [11].

In this paper, we investigate an automatic transcription of ornamented Irish traditional flute music by employing hidden Markov models (HMMs). The proposed system is based on building an individual HMM for each note as well as for each ornament. This enables to model the differences in realisation of ornaments and notes and then detect ornaments whose fundamental frequency is close to the ornamented note. Music signal is represented as a sequence of Mel-frequency cepstral coefficients. Different parameter setups at various stages of the feature extraction and acoustic modelling are explored. Experimental evaluations are performed using recordings of Irish traditional tunes played by flute from Grey Larsen's CD [13]. Evaluations are presented for the task of onset detection and note transcription. Results are presented in terms of the precision, recall and F-measure. Onset detection evaluations are also compared to energy- and F_0-based conventional onset detection algorithms. It is demonstrated that the proposed HMM-based transcription system achieves over 28% relative improvement in terms of the F-measure in onset detection task over conventional onset detection algorithms.

2. ORNAMENTED FLUTE MUSIC

2.1 Ornamentation in Irish traditional flute playing

Ornaments are used as embellishments in Irish traditional music [13]. They are notes of a very short duration, created through the use of special fingered articulations. They can be split into single- and multi-note ornaments. Single-note ornaments, namely 'cut' and 'strike', are pitch articulations. The 'cut' involves quickly lifting and replacing a finger from a tonehole, and corresponds to a higher note than the ornamented note. The 'strike' is performed by momentarily closing an open hole, and corresponds to a lower note than the ornamented note. Multi-note ornaments, namely 'crann', 'roll' and 'shake', are successive use of single-note ornaments. A schematic visualisation of the single- and multi-note ornaments in the time-frequency plane is given in Figure 1.



(a) (b)

Figure 1. A schematic representation of single-note (a) and multi-note (b) ornaments in the time-frequency plane.

2.2 Annotation of the audio data

The audio signal was manually annotated by an experienced player of Irish traditional flute. The annotation provides segmentation of the audio signal, where each segment is characterised by the following: the time of onset, time of offset, type of segment, note identity (if applicable), and note frequency (if applicable). The type of segment may be one of the following: note, one of the types of single-note or multi-note ornaments, and breath. The note

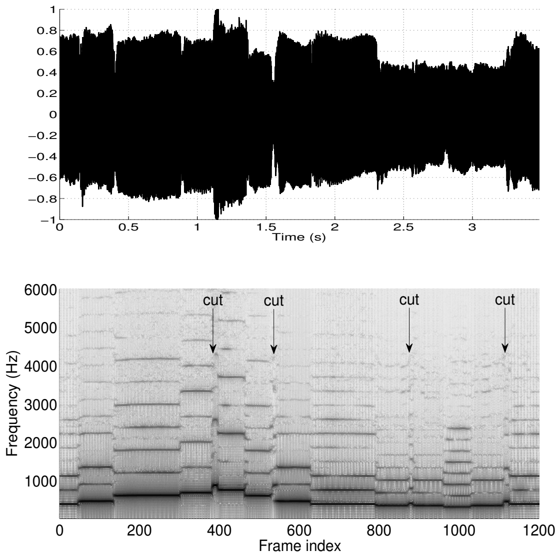


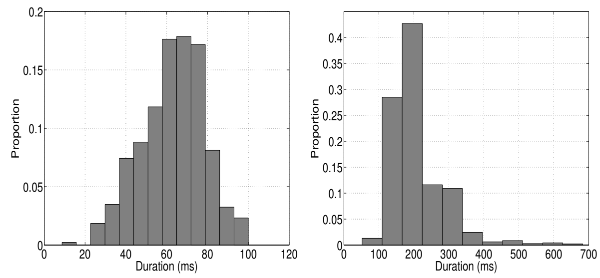
Figure 2. An extract from the tune 'The Lonesome Jig', depicting the waveform (top) and spectrogram (bottom).

frequency is initially estimated automatically but checked by the annotator, and, if needed, then corrected manually. Further information on the process of annotation of the audio recordings is presented in [12].

2.3 Data statistics

The flute music we are dealing with contains notes in the range from D4 to B5, i.e., with the fundamental frequency from 293 Hz to 987 Hz. Typically, only first few harmonics of the notes are having sufficient energy. An example of waveform and spectrogram is given in Figure 2. There are four instances of the 'cut' ornament indicated in the spectrogram at around frame indices 400, 540, 890 and 1130.

Based on the manual annotation, we examined the duration of the notes and single-note ornaments in our recordings. The obtained histograms, depicted in Figure 3, indicate that the duration of ornaments is considerably lower than that of notes. The mean duration of single-note ornaments is 63 ms, while it is 209 ms for notes. In 95% of cases, the duration of single-note ornaments is between 32 ms to 95 ms and of notes between 118 ms to 400 ms.



(a) (b)

Figure 3. The distribution of the duration of single-note ornaments (a) and notes (b) in our recordings.

3. HMM-BASED NOTE TRANSCRIPTION AND ONSET DETECTION

This section presents the proposed HMM-based system for transcription of notes and detection of their onsets. We first describe the representation of the audio signal and then modelling using HMMs.

3.1 Feature representation

The acoustic signal is represented by a sequence of feature vectors, each vector capturing short-time spectral properties of the signal. Since we are dealing with unaccompanied music and in order to obtain lower-dimensional and less-correlated features, the signal is represented using Mel-frequency cepstral coefficients (MFCCs). MFCCs have been widely used in speech and audio processing. The steps involved in converting audio signal into MFCCs is depicted in Figure 4.

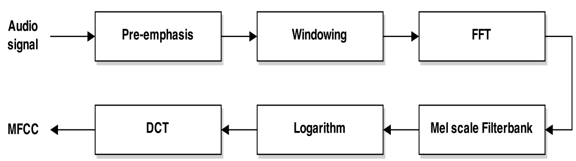


Figure 4. Processing steps used for converting the audio signal into a sequence of Mel-frequency cepstral features.

The signal is first segmented into overlapping frames. The frame length determines the temporal and frequency resolution. While longer frames allow for a finer frequency resolution, we are limited in the setting due to possibly very short duration of ornaments. Each signal frame is multiplied by Hamming window. The windowed frames are then zero padded and the Fourier transform is applied to provide the short-term Fourier spectrum. The short-time magnitude spectrum is passed through Mel-scale filter bank analysis and discrete cosine transform is applied on the logarithm of the filter-bank energies to provide MFCCs. In order to include information on local dynamics, MFCCs were appended with their temporal derivatives, referred to as the delta and accelleration features, which were calculated as in [17]. The values of the parameters within the processing steps need to reflect that we are dealing with ornamented flute music. As such, in our experimental evaluations, we explored various parameter setups.

3.2 Modelling

The model of each note is obtained by modelling the temporal evolution of feature vectors using a left-to-right (no skip allowed) hidden Markov model (HMM). We found that in addition to having an HMM for each note, it is essential to have also a separate HMM for each single-note ornament. This allows to deal with the fact that the realisation of single-note ornaments may not fully reach the notional frequency of a note but rather be somewhere between two notes. In addition to this, we also create a model for breath and silence. These are used to model the taking a breath by the player and the initial and final silences in recordings, respectively. Overall, we have 42 models, consisting of 14 models for notes, 14 models for cuts, 11 models for strikes (strikes for some notes did not occur in our data) plus breath and silence. The state output probability density function (pdf) at each HMM state is modelled using a mixture of Gaussians. This allows for a more accurate modelling of variations in playing notes than using a single Gaussian distribution. Gaussian distributions with a diagonal covariance matrix are used due to computational reasons, as is typically done in speech and audio pattern processing. We explore the effect of using different number of HMM emitting states and Gaussian mixture components in the experimental section. The transcription system was built using the HTK [17].

3.2.1 *Training*

The initial values for the parameters of individual HMMs were estimated using isolated extracts from the audio signal, by applying several iterations of the Viterbi style training procedure. The isolated extracts were obtained based on the manual time-stamp annotation, i.e., onset and offset times. Further training of the models was then performed using several iterations of the Baum-Welch (aka forward-backward) algorithm. This uses continuous audio signal as input and requires only the sequence of notes/ornaments labels (i.e., no time-stamp). As such, this can eliminate the effect of possible errors in time-stamp annotation at borders of notes/ornaments on the trained models.

3.2.2 *Recognition*

To perform recognition, we need to construct a recognition network. This defines the allowed sequences of models (i.e., notes/ornaments). A network that closely reflects the knowledge of music could be employed. In this paper, we did not employ any such knowledge. We used a loop network that allows any note to follow any other note. We modified this slightly to reflect that an ornament model need to be followed by a note model. The network we used is depicted in Figure 5. As this network allows the same note to be subsequently repeated in the recognition output, we post-process the results such that the repetitions of the same note are considered to be a single instance of the note, for example, the original recognition output B4 D4 D4 A5 is considered to be B4 D4 A5. Note that a fixed probability value, aka word insertion penalty, can be associated with the transition from the end of one model to the start of the next model. This is useful in controlling the balance between the number of models being incorrectly inserted and deleted in the recognition result and we used it in our experimental evaluations.

Given a sequence of feature vectors, the Viterbi algorithm is used to find the optimal state sequence. This provides the sequence of recognised models as well as the start (and the end) times of each recognised model, i.e., the onset detection result.

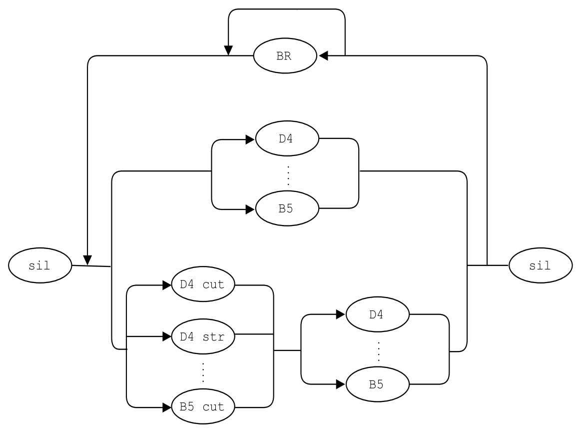


Figure 5. The recognition network used in the HMM-based note transcription system. The elipses denote individual HMMs. Models of ornaments are denoted by note identity appended with either 'cut' or 'str', representing models for 'cut' and 'strike' ornaments, respectively. 'BR' denotes breath and 'SIL' silence models.

4. EXPERIMENTAL EVALUATIONS

4.1 Data description

Evaluations are performed using recordings of Irish traditional tunes and training exercises played by flute from Grey Larsen's CD which accompanied his book "Essential Guide to Irish Flute and Tin Whistle" [13]. The tunes are between 16 sec and 1 min 22 sec long. All recordings are monophonic and are sampled at 44.1 kHz sampling frequency.

The collection consists of 19 tunes. The list of the tunes, with the number of notes and ornaments, is given in Table 1. In total, there are 3929 onsets, including notes and ornaments. Out of these there are 804 single-note ornaments (which includes also counts from parts of multi-note ornaments), consisting of 620 cuts and 184 strikes.

First evaluations are performed to demonstrate the effect of different parameter setups – due to computational reasons, these experiments use all files for training of models and also for testing. Final evaluations are performed using the leave-one-out cross-validation procedure, in which in turn 1 file is kept for testing and all the 18 remaining files are used for training. The results were accummulated over all files and then the evaluation measures calculated.

4.2 Evaluation measures

Performance of both the onset detection and note recognition is evaluated in terms of the precision (P), recall (R) and F-measure. The definition of these measures is the same as used in MIREX onset detection evaluations, specifically,

$$P = \frac{N_{tp}}{N_{tp} + N_{fp}}, R = \frac{N_{tp}}{N_{tp} + N_{fn}}, F = \frac{2PR}{P + R}.$$

In the case of onset detection, N_{tp} is the number of correctly detected onsets and N_{fp} and N_{fn} is the number of inserted and deleted onsets, respectively. The onset detection is considered as correct when it is within ± 50 ms around the onset annotation.

In the case of note recognition, N_{tp} is the number of correctly recognised notes and N_{fp} and N_{fn} is the number of inserted and deleted notes, respectively.

Tune Title	Number of			Time
	Notes	Ornaments		(sec.)
		Cut	Strike	
Study 5	55	16	0	20
Study 6	56	24	0	20
Study 11	76	20	0	26
Study 17	48	19	0	16
Study 22	127	0	28	47
Maids of Ardagh	98	28	5	32
Hardiman the ..	112	22	7	28
The Whinny Hills ..	117	34	6	30
The Frost is All ..	151	39	14	41
The Humours of ..	289	113	19	82
The Rose in the ..	152	33	13	39
Scotsman over ..	153	33	9	38
A Fig for a Kiss	105	27	9	28
Roaring Mary	176	42	22	44
The Mountain Road	105	20	6	25
The Shaskeen	181	52	23	42
Lady On The Island	118	21	1	21
The Lonesome Jig	153	27	0	46
The Drunken ..	185	50	22	43
Total	2457	620	184	668

Table 1. The list of tunes contained in the dataset, with the number of onsets and single-note 'cut' and 'strike' ornaments and duration of each tune.

4.3 Results for various parameter setups in feature extraction and modelling

This section explores the effect of different setups of parameters in the feature extraction and HMM-based modelling on the task of onset detection. A comparison with three conventional onset detection methods is also given.

4.3.1 Conventional onset detection algorithms

The conventional onset detection methods we employed to provide a comparison are: two methods which exploit the change of the signal amplitude over time, with processing performed in the temporal and spectral domain, and a method based on the fundamental frequency (F_0). The description of these methods, which we also used in our previous onset detection research, is provided in [11].

We performed extensive evaluations with different parameter values for each of the conventional onset detection methods. The best achieved performance for each of the methods is presented in Table 2. It can be seen that the F_0-

based method performed better than each of the energy-based methods and achieved the F-measure of 91.2%.

Algorithm for onset detection	Evaluation performance (%)		
	Precision	Recall	F-measure
sig-energy (spectral)	94.8	85.6	89.9
sig-energy (temporal)	90.8	88.4	89.6
F_0	89.3	93.2	91.2

Table 2. Results of onset detection obtained using conventional onset detection methods.

4.3.2 HMM-based onset detection

Now we explore the performance of the HMM-based system when using various parameter setup in the feature extraction and acoustic modelling.

First, we compare results achieved by the HMM-based system when using the estimated F_0 with energy and the MFCCs (see the first row in Table 4 for the parameter setup) as the input features. Results are presented in Table 3. It can be seen that when using the estimated F_0 as input features to HMMs, the F-measure improved to 93.8%, from 91.2% that was achieved using the conventional F_0-based method (as in Table 2). The performance of the HMM-based system improved considerably further to 96.7% when using MFCC features as input, instead of the estimated F_0. As such, the use of HMMs driven with MFCC features provided over 60% error rate reduction over the best conventional method. The considerably better performance of the HMM-based system may be attributed to several factors. First, it is the statistical modelling of the temporal evolution of features. Second, the features used provide information about the spectral content. This is unlike the energy-based and F_0-based conventional methods which accummulate the information from the entire signal bandwidth into a single detection function or into an F_0 estimate. Third, the use of HMM effectively incorporates smoothing of the frame-based decisions and imposes a minimum duration of notes and ornaments.

Features input to HMM	F-measure (%)
F_0 and energy, both with Δ and Δ^2	93.8
MFCC, both with Δ and Δ^2	96.7

Table 3. Results of the HMM-based onset detection when using an estimate of F_0 and MFCCs as input features.

Results obtained using diferent parameter setups in the MFCC feature extraction are presented in Table 4. The first line in the table presents the best parameter setup values and this is: bandwidth of 4 kHz, frame length of 12.5 ms, frame-shift of 5 ms, Mel-scale filter-bank with 21 channels, using 12 cepstral coefficients, and appending delta and acceleration coefficients (which were extracted using

Parameters in MFCC feature extraction		F-measure (%)
BW=4kHz, FrmL=12.5ms, FrmS=5ms, nFB=21, nCC=12, Δ and Δ^2		96.7
Bandwidth	6 kHz	95.5
(BW)	8 kHz	95.4
Frame-length	10 ms	95.9
(FrmL)	15 ms	96.3
	20 ms	96.3
	30 ms	95.7
Frame-shift	3 ms	95.5
(FrmS)	7 ms	95.4
number of Mel filter-bank	17	96.1
(nFB)	25	96.5
Cepstral coefficients	10	95.8
(nCC)	14	96.6
Δ^2 coefficients	no	95.8
Δ and Δ^2 coefficients	no	91.6

Table 4. Results of the HMM-based onset detection in terms of the F-measure obtained with different parameter setup in MFCC feature extraction.

the window of 3 and 2 signal frames, respectively). We now analyse the effect of each parameter – in each experiment, only one parameter is modified at a time in reference to the above best parameter setup. Let us start with varying the frequency bandwidth of the signal. This was performed at the stage of designing the Mel filter-banks. For the bandwidth of 6 kHz and 8 kHz, the number of filters was adjusted such that the lower 4 kHz was in all cases covered by 21 filters. It can be seen that the performance is similar when using the bandwidth of 6 kHz and 8 kHz but it improves considerably when the bandwidth is reduced to 4 kHz. This reflects, as we have also noticed in our visual inspection of spectrograms, that there is little signal content above 4 kHz in our flute recordings and as such the inclusion of the higher frequency range acts detrimentally to performance. This result may be used when analysing flute playing that contains accompaniments in higher frequencies or is recorded in live performances where other unwanted sounds may be present in higher frequencies. Next, results using different length of frames show that similar performance is achieved for lengths between 12 to 20 ms. The performance starts to decrease considerably when frames of 30 ms are used. This is due to the presence of ornaments, duration of which may be as short as 20 ms. In the case of frame shift, it can be seen that setting this to 3 ms or 7 ms considerably degrades the performance in comparison to the use of 5 ms shift. Varying the number of filter-bank channels from 17 to 25 has only relatively little effect, with performance being at the peak for 21 channels. The use of 12 or 14 cepstral coefficients provides very similar performance, while reducing this to 10 has quite negative effect. Finally, experiments when the delta and accelleration features, denoted by Δ and Δ^2, respectively, are not included in feature representation are presented. Results show a large decrease in performance when neither delta nor accelleration features are used. This demonstrates the importance of incorporating information on local dynamics of the acoustic signal.

Next, we present the effect of different parameter setups in acoustic modelling. We vary the number of states and of mixture components of each state pdf for models of notes and ornaments. Results are presented in Table 5. The suitable range of values for the number of states of note and ornament models is determined based on the frame shift used in the feature extraction and statistics of the duration of the notes and ornaments. As such, we explored the range from 6 to 12 for notes and from 2 to 6 for ornaments. It can be seen that there is not much performance variation when using this range of values. In regard to the number of mixture components, it can be seen that it is useful to have at least 4 mixture components for note models, while even 2 mixture components seem sufficient for models of ornaments.

Parameters in acoustic modelling		F-measure (%)
nStates for N / O / B: 8 / 4 / 8, nMix for N / O / B: 6 / 2 / 6		96.7
nStates for notes	6	95.9
	10	96.1
	12	95.8
nStates for ornaments	2	95.8
	6	96.4
nMix for notes	2	95.6
	4	96.3
	8	96.6
nMix for ornaments	1	96.0
	4	96.7
	6	96.6

Table 5. Results of HMM-based onset detection in terms of the F-measure obtained with different parameter setup in acoustic modelling. N, O, and B stand for note, ornament and breath, respectively.

4.4 Results using the leave-one-out cross-validation

The final experimental evaluations are performed using the leave-one-out cross-validation. The feature extraction and acoustic modelling parameter setup that achieved best performance in previous section is used. The achieved results of onset detection and note identity recognition are presented in Table 6. It can be seen that very good performance is obtained for both tasks. The drop in onset detection performance in comparison to the results presented in the previous section is expected as the testing files have now not been seen during the training. Nevertheles, the performance is improved by over 28% relative over the conventional onset detection algorithms whose parameters were actually tuned based on both training and testing data.

	Evaluation performance (%)		
	Precision	Recall	F-measure
Onset detection	95.0	92.4	93.7
Note recognition	96.4	95.2	95.8

Table 6. Results of onset detection and note recognition obtained by the HMM-based system using the leave-one-out cross-validation procedure.

5. CONCLUSION AND FUTURE WORK

In this paper, we presented work on transcription of Irish traditional flute music containing ornamentation. The presented system is based on modelling of individual notes and ornaments using hidden Markov models. Acoustic signal is represented as a sequence of Mel frequency cepstral coefficients. A wide range of parameter setup values in both the feature extraction and acoustic modelling were explored. Experimental evaluations were performed using recordings of 19 Irish traditional flute tunes, containing in total 3929 onsets, out of which 804 corresponds to ornaments. Using the leave-one-out cross-validation procedure, the proposed HMM-based system achieved the F-measure of 93.7% in detecting onsets of ornaments and notes. This represents over 28% error rate reduction compared to conventional onset detectors whose parameters were even tuned to the testing data. The F-measure in the task of recognising the note identity was 95.8%.

There are several possible extensions of this work we are currently considering. First, the presented evaluations were performed using recordings from the same CD. We plan to perform evaluations on a range of recordings from several CDs in order to explore the capability of the system in dealing with variability due to different recording conditions, makes of flute instruments and performers. We will investigate techniques to compensate for such variability in order to improve robustness. Second, we plan to analyse the errors the automatic system makes in onset detection and note identity recognition tasks and reflect this in modifications to the system to further improve the performance. Then, the current HMM-based framework allows directly and in a probabilistic manner to incorporate musical knowledge on the sequences of notes used in flute music. Such knowledge could be obtained from musicologists and/or extracted automatically from annotations. Next, incorporation of an explicit duration modelling of notes and ornaments could help to reduce the number of falsely inserted and deleted onsets. Finally, we plan to expand the system to deal with recordings, in which the flute is accompanied by other instruments and/or singing.

6. ACKNOWLEDGEMENTS

This work was supported by the project ‘Characterising Stylistic Interpretations through Automated Analysis of Ornamentation in Irish Traditional Music Recordings’ under the ‘Transforming Musicology’ programme funded by the Arts and Humanities Research Council (UK).

7. REFERENCES

[1] J. P. Bello, G. Monti, and M. Sandler. Techniques for automatic music transcription. In *Int. Symposium on Music Information Retrieval (ISMIR)*, Plymouth, USA, 2000.

[2] E. Benetos and S. Dixon. Joint multi-pitch detection using harmonic envelope estimation for polyphonic music transcription. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1111–1123, October 2011.

[3] A.T. Cemgil, H.J. Kappen, and D. Barber. A generative model for music transcription. *IEEE Trans. onAudio, Speech, and Language Processing*, 14(2):679–694, March 2006.

[4] A. de Cheveigne and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111(4):1917–1930, April 2002.

[5] K. Dressler. Extraction of the melody pitch contour from polyphonic audio. In *Int. Conf. on Music Information Retrieval (ISMIR)*, London, UK, 2005.

[6] V. Emiya, R. Badeau, and B. David. Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches. In *16th European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, 2008.

[7] B. Fuentes, R. Badeau, and G. Richard. Harmonic adaptive latent component analysis of audio and application to music transcription. *IEEE Trans. on Audio, Speech, and Language Processing*, 21(9):1854–1866, 2013.

[8] M. Gainza and E. Coyle. Automating ornamentation transcription. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, 2007.

[9] S. Godsill and Manuel Davy. Bayesian harmonic models for musical pitch estimation and analysis. In *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages II–1769–II–1772, May 2002.

[10] A.P. Klapuri. Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. *IEEE Transactions on Speech and Audio Processing*, 11(6):804–816, Nov 2003.

[11] M. Köküer, P. Jančovič, I. Ali-MacLachlan, and C. Athwal. Automatic detection of single and multi-note ornaments in Irish traditional flute playing. In *Proc. of the15th Int. Society for Music Information Retrieval Conf. (ISMIR)*, pages 15–20, Taipei, Taiwan, Nov 2014.

[12] M. Köküer, D. Kearney, I. Ali-MacLachlan, P. Jančovič, and C. Athwal. Towards the creation of digital library content to study aspects of style in Irish traditional music. In *Proc. of the Int. Workshop on Digital Libraries for Musicology (DLFM)*, London, 2014.

[13] G. Larsen. *The Essential Guide to Irish Flute and Tin Whistle*. Mel Bay Publications, Pacific, Missouri, USA, 2003.

[14] G. E. Poliner and D. Ellis. A discriminative model for polyphonic piano transcription. *EURASIP Journal on Advances in Signal Processing*, 2007(1):048317, 2007.

[15] M. P. Ryynänen and A. P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32(3):72–86, 2008.

[16] J. Salamon and E. Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Trans. on Audio, Speech, and Language Processing*, 20(6):1759–1770, 2012.

[17] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland. *The HTK Book. V2.2*. 1999.