Player recognition for traditional Irish flute recordings using K-nearest neighbour classification

Islah Ali-MacLachlan[1†], Edmund Hunt[2] and Alastair Jamieson[1],
[1] DMTLab, Birmingham City University, Birmingham B4 7XG, UK
[2] Royal Birmingham Conservatoire, Birmingham B4 7XR, UK
[†] Corresponding author: islah.ali-maclachlan@bcu.ac.uk

Introduction

Irish traditional music (ITM) is central to the idea of Irish cultural identity. The wooden traverse flute is one of several instruments used in ITM alongside the fiddle (violin), uilleann pipes and tin whistle. Mastery in musicianship is displayed by a player's translation of an unadorned traditional melody into a personalised rendition containing stylistic traits such as ornamentation, dynamics and timbre (Breathnach, 1996). In ITM, stylistic differences between flute players have been attributed to many influences, including regional playing styles, *sean nós* singing, uilleann pipes technique, and players' personal preferences (Johnston, 1995). Within the ITM community, the importance of timbre is illustrated by the frequent use of adjectives such as 'reedy', 'earthy' or 'warm' to describe an individual flute player's tone. However, these descriptions of timbre are highly subjective and difficult to analyse.

The aim of this work is to identify individual traditional Irish flute players from recordings, and to understand the influence of two types of features, harmonic magnitudes and and mel-frequency cepstral coefficients (MFCCs) in attaining an overall classification accuracy. A number of experiments have used differences between magnitudes of a range of harmonics along with changes in amplitude envelope to indicate timbral variances (Grey, 1977; Iverson & Krumhansl, 1993). Harmonic magnitudes can be used to identify individual flute players of different ability levels (Ali-MacLachlan et al., 2013). MFCCs are used to discriminate between sonorant and non-sonorant speech (Dumpala et al., 2015) and have been used successfully in flute player identification (Ali-MacLachlan et al., 2018). Studies have shown that it is difficult to identify professional flute players from recordings by using only harmonic magnitudes, due to experienced players having more breath and embouchure control (Ali-MacLachlan et al., 2015). Studies often use the steady-state central part of notes and in this work we also investigate a comparison between steady-state, attack, release and whole event.

Method

The recordings chosen for analysis were part of the ITM-Flute-99 dataset[1]. We used the dataset described in (Ali-MacLachlan et al., 2015) containing five players, each playing four solo unaccompanied traditional Irish flute tunes. All of these players are regarded as having a distinctive musical signature. A typical playing style includes the ornamentation of a traditional melody with *cuts* and *taps* – very short pitch deflections up and down caused by quick finger movements over the holes above and below. These ornaments are usually in the range of 0.02-0.08 sec. whereas most melody notes are in the range of 0.1-0.3 sec. (Köküer et al., 2014).

The audio for each event is extracted by using ground truth timing data and in the case of shorter events, zero padded to 2048 samples. The Fourier transform is applied and the absolute values are taken to provide a spectrum. Magnitudes are then compressed by applying logarithm. H1 – H5 harmonic magnitudes are identified semi-automatically by calculating the localised maxima around the multiples of the event frequency from 1 to 5. 13 MFCC coefficients are also extracted and the first one is discarded as it contains a constant offset relating to the average log-energy of the input signal.

A study was conducted using a KNN classifier with a range of k=1-10 for a dataset using all notes (see Table 1). Based on this, k=1 was found to return the highest accuracy.

[1] www.github.com/izzymaclachlan/datasets

Table 1. Comparison of identification accuracy (%) for the Notes dataset using different k values with HMAG and MFCC features

k	1	2	3	4	5	6	7	8	9	10
Notes	91.2	88.9	90.3	88.5	88.5	86.9	87.3	86.2	86.3	85.8

The KNN classifier (k=1) was then trained with harmonic magnitude (HMAG) and mel frequency cepstral coefficient (MFCC) features derived from attack, sustain or release portions of events. 4-fold cross-validation was used, where 75% of the data was combined to train the model, which was tested on the remaining 25%. The process was repeated four times and overall performance aggregated across the four folds.

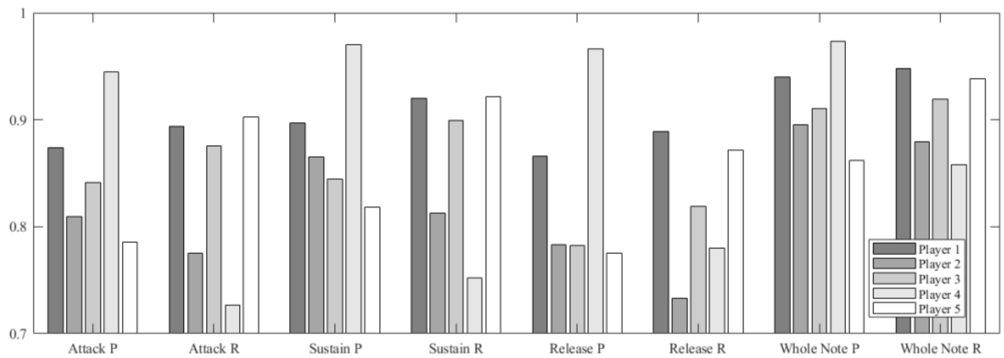

Figure 1- Comparison of precision (P) and recall (R) for attack, sustain, release and whole notes, showing results per class. All HMAG and MFCC features are used on audio of notes only.

Results

Figure 1 shows precision and recall results across 5 players when the classifier is trained using data from either attack, sustain (steady state) or release sections, or by using the whole note. The average precision and recall for the sustain section is 0.8148 and 0.7667 in comparison to whole notes at 0.9162 and 0.9087. The precision and recall levels for different players are variable showing that different styles contribute to individuality in different sections of the note.

Figure 2 shows that harmonic magnitudes make very little contribution towards precision and recall. Average precision and recall across 5 players is 0.9164 and 0.9073 when using MFCC only, and 0.9162 and 0.9087 when using both MFCC and HMAG. Using higher order harmonics H4 and H5 alongside H1, H2 and H3 increase classification accuracy when HMAG is used alone.

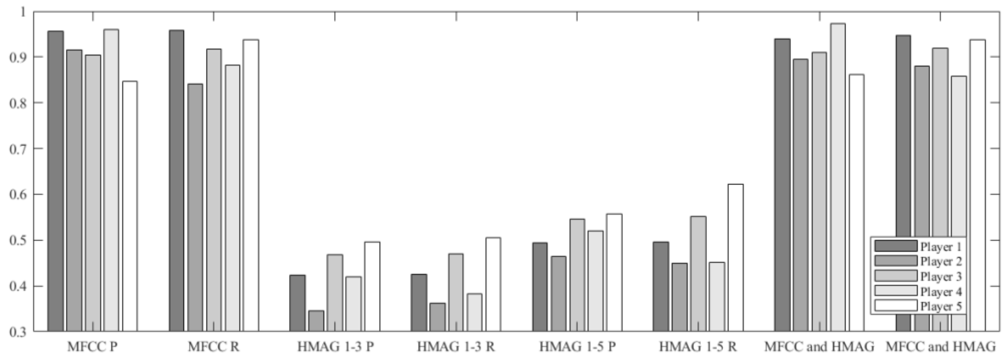

Figure 2- Comparison of precision (P) and recall (R) for MFCC only, HMAG H1-H3, HMAG H1-H5 and MFCC with HMAG H1-H5. All HMAG and MFCC features are used on audio of notes only.

Figure 3 shows a differences in precision and recall when training a model with notes or ornaments individually or together (all events). Average precision and recall across 5 players is 0.9164 and 0.9073 with notes only in comparison to 0.8569 and 0.8519 when training with all events.

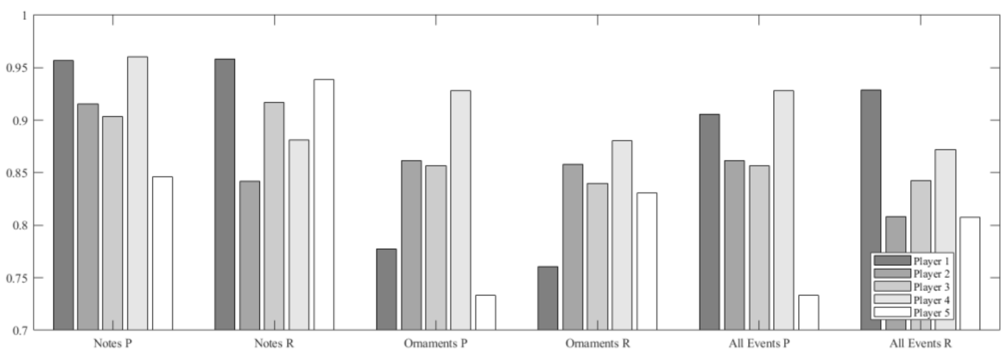

Figure 3- Comparison of precision (P) and recall (R) for notes, ornaments and all events (notes + ornaments). MFCC features are used in all cases.

Conclusions

The findings contribute to a better understanding of timbre in traditional flute playing by allowing us to quantify the difference in results that are achieved by analysis of different note sections and use of harmonic magnitudes in comparison to MFCCs. The results show that MFCCs return substantially better results and that using harmonic magnitudes alongside make little difference.

Our previous studies have concentrated on using the steady state central section of notes but this study shows that the average precision and recall across the 5 player classes for whole notes is higher than for the central section only. We also found that using longer melodic notes without shorter ornaments give better results.

In future research, we hope to compare these results to listening tests in order to explore the correlation between computational analysis and a flute player's comprehension of style and tonal quality.

References

Ali-MacLachlan, I., Köküer, M., Athwal, C., & Jancovic, P. (2015). Towards the identification of Irish traditional flute players from commercial recordings. In: Association Dirac (ed), *Proceedings of the 5th International Workshop on Folk Music Analysis (FMA) (pp.13-17)*. Paris, France.

Ali-MacLachlan, I., Köküer, M., Jancovic, P., Williams, I., & Athwal, C. (2013). Quantifying Timbral Variations in Traditional Irish Flute Playing. In: Kranenburg P., Anagnostopoulou C., Volk A. (eds), *Proceedings of the 3rd International Workshop on Folk Music Analysis (FMA)*. (pp. 7–14). Amsterdam, Netherlands.

Ali-MacLachlan, I., Southall, C., Tomczak, M., & Hockman, J. (2018). Player recognition for traditional Irish flute recordings. In: Holzapfel A., Pikrakis A. (eds), *Proceedings of the 8th International Workshop on Folk Music Analysis (FMA)* (pp.3-7). Thessaloniki, Greece.

Breathnach, B. (1996). *Folk music and dances of Ireland*. London: Ossian Publications Ltd.

Cowdery, J. R. (1990). *The melodic tradition of Ireland*. Ohio: Kent State University Press.

Dumpala, S. H., Nellore, B. T., Nevali, R. R., Gangashetty, S. V., & Yegnanarayana, B. (2015). Robust features for sonorant segmentation in continuous speech. In: Möller S., Ney H., Möbius B., Nöth E., Steidl S. (eds), *Sixteenth Annual Conference of the International Speech Communication Association*. (pp.1987–1991). Dresden, Germany.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*(5), 1270–1277.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *The Journal of the Acoustical Society of America*, *94*(5), 2595.

Johnston, T. F. (1995). The Social Context of Irish Folk Instruments. *International Review of the Aesthetics and Sociology of Music*, *26*(1), 35–59. https://doi.org/10.2307/836964

Köküer, M., Ali-MacLachlan, I., Jancovic, P., & Athwal, C. (2014). Automated Detection of Single-Note Ornaments in Irish Traditional flute Playing. *Proceedings of the 4th International Workshop on Folk Music Analysis (FMA)*. Istanbul, Turkey.

Widholm, G., Linortner, R., Kausel, W., & Bertsch, M. (2001). Silver, gold, platinum-and the sound of the flute. In: Bonsi D. (ed), *Proceedings of the International Symposium on Musical Acoustics*, *1*, 277–280.