

PLAYER IDENTIFICATION FOR TRADITIONAL IRISH FLUTE RECORDINGS

First Author

Affiliation1

author1@ismir.edu

Second Author

Retain these fake authors in

submission to preserve the formatting

Third Author

Affiliation3

author3@ismir.edu

ABSTRACT

The abstract should be placed at the top left column and should contain about 150-200 words.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

1. INTRODUCTION

Irish Traditional Music (ITM) has progressed from its roots of social dance music to become a medium for listening. Represented at several large-scale music festivals in Europe and the US, and over 1500 weekly 'sessions' where musicians can meet and share repertoire, the indigenous music of Ireland has a large, international audience [30].

Playing of the wooden simple system flute in ITM was historically linked predominantly to the west and north-

west of Ireland. The Irish cultural revival has allowed the flute to play a wider role as an intrinsic part of the sound of ITM alongside tin whistles, accordions, and fiddles [33].

Traditional flute players are individuated based on their use of techniques such as ornamentation, phrasing and articulation [14, 15, 17, 24] alongside idiosyncratic timbral differences [1, 2, 32]. Mastery is judged by creativity and technical ability within the stylistic bounds of traditional music and distinctive characteristics often develop organically as a player learns and refines their skills.

In order to determine stylistic differences between players, we must first develop methods to detect these differences in audio signals. We evaluate recordings produced by six accomplished traditional flute players, all playing music from a predetermined corpus offering a range of typical modes and rhythms.

1.1 Related work

Automatic music genre classification was initially introduced by Tzanetakis & Cook [29] who proposed a framework for the development and analysis of features for music content analysis. As part of this work, mel-frequency cepstral coefficients (MFCC) features [22] were implemented due to their efficient representation of spectral data. Since then, several authors have proposed systems for genre and artist identification.

Li and Ogihara [20] implemented a semi-supervised learning system with timbral features such as MFCC, spectral centroid, rolloff and flux as input alongside lyrical content. Mandel & Ellis [23] used support vector machines (SVM) to identify a single popular music performer from a group of 18 in the *uspop2002* corpus [5, 12]. Input features were segmented 20 band MFCC extracted using the method described in [26]. An important feature of this study was to explore and negate the 'album effect' [16, 31] where songs from a single album are spectrally more similar than songs from different albums.

State of the art identification techniques have been used as part of automatic music classification. Convolutional neural networks (CNN) have been used successfully with input features derived from spectrograms [18] and MFCC representing timbre, tempo and key variations [19]. A CNN was trained to perform artist and genre recognition on the 'Million Song Dataset' [6] using segments related to note onsets and feature vectors containing timbre and chroma components [11]. Lidy & Schindler [21] used CNN to classify genre, mood and composer and achieved



© First Author, Second Author, Third Author. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** First Author, Second Author, Third Author. "PLAYER IDENTIFICATION FOR TRADITIONAL IRISH FLUTE RECORDINGS", 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

the highest results in MIREX 2016 using a 40-band Mel filter.

Costa et al [10] reinforced the effectiveness of CNN for music genre classification, performing analysis against SVM on three separate databases containing Western, Latin and African music. The CNN system compared favourably in a number of scenarios, particularly in the case of African music where the sources were largely field recordings. This method showed favourable results with multiple class types labelled country, function, ethnic group and instrumentation. Individual instrument classification is discussed in [27] and as part of an ensemble in [13].

Studies in flute acoustics have found that individual players produce markedly different timbres while changes in material make very small spectral differences [4, 9, 32]. Previous methods of player detection in ITM have used signal processing methods [1, 2]. The use of probabilistic modelling, in particular CNN, is an important step in more accurate player identification. We propose a CNN system for identification of flute players in ITM in order to make use of their ability to efficiently process large datasets.

The remainder of this paper is structured as follows: Section 2 details CNN and their implementation. In Section 3 we discuss evaluation of.... Results of the studies into ... are presented in Section 4 and finally conclusions and further work are discussed in Section 5.

2. METHOD

Introductory paragraph

2.1 Convolutional neural networks (CNN)

In order to capture the stylistic differences between each player we divide recordings into 5 second segments. We then extract timbral features as these are important in class distinction. Each player contributes distinct timbral features as part of their style. For that purpose we extract 40 MFCCs, excluding the first coefficient, to accommodate for differences in timbre over a range of played notes.

CNNs enable larger input feature sizes to be used by first reducing the dimensionality of the features using a shared weighted kernel and pooling before they are input into a fully connected layer. This ability has enabled CNNs to achieve higher accuracies than recurrent neural networks in closed related fields including onset detection, beat detection and downbeat detection. REFERENCES An overview of the implemented CNN ADT systems with different input feature sizes is outlined in Figure... REF. All of the implement versions consist of two sets of convolutional, max pooling, dropout, and batch normalisation layers before a 100 neuron fully connected layer and a two neuron softmax output layer. The output h of a two-dimensional convolutional layer with a rectified linear unit transfer function is calculated using:

$$h_{ij}^f = r \left(\sum_{l=0}^{L-1} \sum_{m=0}^{M-1} W_{ml}^f x_{(iu+l)(jv+m)} + b^f \right) \quad (1)$$

where x is the input features, W and b are the shared weights and bias and f is the feature map. L and M are the dimensions of the shared weight matrix and I and J are the output dimensions of that layer. u and v are the stride lengths and the equation for the rectifier linear unit transfer function r is:

$$r(\phi) = \max(0, \phi) \quad (2)$$

Before the features x are input into the network they are zero padded.

The output of the convolutional layer h is then processed using a max pooling layer where the maximum of a by b sub-regions is taken. Dropouts are then performed on the output of the max pooling layer with a dropout probability of 0.25 and batch normalisation is implemented. **IN DEPTH DETAIL ON BATCH NORM!**

The output of the second batch normalisation h_{bnL} layer is then input into a dense fully-connected layer ($Q = h_{bnL}$) Equation.?? with a rectifier linear unit transfer function ($\sigma(\phi) = r(\phi)$) Equation.2. The output layer of the implemented CNNs is the same as the output layer of the SABRNN systems with the input to the layer being the output of the fully connected layer ($Q = h_{fc}$).

2.2 Implementation

2.2.1 Input Features

2.2.2 Training Procedure

The implemented models are trained using the Adam optimizer and a learning rate of 0.003. Mini-batch gradient descent is used with batch sizes of 250. Training is stopped when a minimum of 50 iterations have commenced and the validation set accuracy has not increased between iterations. The weights are initialized using a scaled uniform distribution [28] and biases are initialized to 0. Cross entropy is used as the loss function.

3. EVALUATION

3.1 Dataset

For these evaluations, we require a dataset that is representative of a range of respected players with individual stylistic traits. The dataset, containing 247 recordings of traditional 'tunes' played on solo flute, includes two discrete sections. The first section¹ comprises of 28 recordings by each of 6 experienced players (168 total) and the second section includes 79 released recordings of 9 professional players detailed in [3]. Only the first section was used in evaluation 1 (see Section 4) whereas sections 1 and 2 were both used in evaluation 2.

The dataset is representative of the four most popular melodic styles, namely the reel, jig, hornpipe and polka [14]. The structure of most traditional melodies in ITM can be attributed to two scales, D and G, and four classes

¹ recordings used in this section are available at: <https://github.com/izzymaclachlan/datasets>

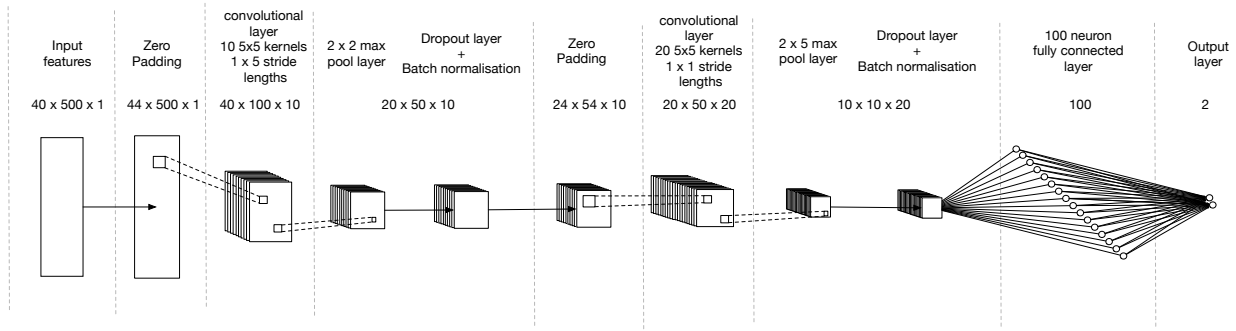


Figure 1. Overview of the proposed implemented CNN system.

Tune Title	Type	Scale	Ends on
Maids of Mount Cisco	Reel	G	Ray
The Banshee	Reel	G	Soh
Cooley's Reel	Reel	G	Lah
Banish Misfortune	Jig	G	Doh
Morrison's Jig	Jig	D	Ray
The Home Ruler	Hornpipe	D	Doh

Table 1. Corpus recorded by all players detailing tune type, scale and ending note

defined by the ending note of the tune and given the solfa name Doh, Ray, Soh and Lah [8, 25].

The corpus of tunes shown in Table 3.1 was chosen to be representative of a typical player repertoire and to represent a range of scales and classes. 6 players played each tune twice without tempo restriction and twice in time with a click track played through headphones and set to represent a typical tempo according to [7]. Players also recorded two other tunes picked from their own repertoire.

3.1.1 Recording

The recordings were collected as 16-bit/44.1kHz WAV files using a Thomann MM-1 measurement microphone connected to an Audient ID14 audio interface. The microphone was positioned above the middle of the flute in order to minimise wind noise.

3.2 Evaluation of methods

4. RESULTS

We present two evaluations using the datasets presented in Section 3.1. In the first, we analyse 168 recordings of 6 individual flute players and present results of 1 vs all (2 classes) and 1-6 vs all (6 classes). In the second evaluation we also include an additional 79 recordings of 9 different players, presenting the results of 1 vs all (2 classes) and 1-7 vs all (7 classes).

4.1 Evaluation 1 - 6 Players

Can't make confusion matrixes look right in column at present

4.2 Evaluation 2 - 6 Players and additional dataset

4.3 Sub3

5. CONCLUSIONS AND FUTURE WORK

6. REFERENCES

- [1] I. Ali-MacLachlan, M. Köküer, C. Athwal, and P. Jančovič. Towards the identification of Irish traditional flute players from commercial recordings. In *Proceedings of the 5th International Workshop on Folk Music Analysis*, Paris, France, 2015.
- [2] I. Ali-MacLachlan, M. Köküer, P. Jančovič, I. Williams, and C. Athwal. Quantifying Timbral Variations in Traditional Irish Flute Playing. In *Proceedings of the 3rd International Workshop on Folk Music Analysis*, Amsterdam, Netherlands, 2013.
- [3] I. Ali-MacLachlan, M. Tomczak, C. Southall, and J. Hockman. Note, cut and strike detection for traditional Irish flute recordings. In *Proceedings of the 6th International Workshop on Folk Music Analysis*, Dublin, Ireland, 2016.
- [4] J. Backus. Effect of wall material on the steady-state tone quality of woodwind instruments. *The Journal of the Acoustical Society of America*, 36(10):1881–1887, 1964.
- [5] A. Berenzweig, B. Logan, D. PW Ellis, and B. Whitman. A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*, 28(2):63–76, 2004.
- [6] T. Bertin-Mahieux, D. Ellis, B. Whitman, and Paul Lamere. The million song dataset. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 2011.
- [7] B. Breathnach. *Ceol Rince na hÉireann 1*. An Gum, Dublin, Ireland, 1963.
- [8] B. Breathnach. *Folk music and dances of Ireland*. Osian, London, 1996.
- [9] J.W. Coltman. Effect of material on flute tone quality. *The Journal of the Acoustical Society of America*, 49:520, 1971.

- [10] Y. Costa, L.S. Oliveira, and C.N. Silla. An evaluation of convolutional neural networks for music classification using spectrograms. *Applied Soft Computing*, 52:28–38, 2017.
- [11] S. Dieleman, P. Brakel, and B. Schrauwen. Audio-based music classification with a pretrained convolutional network. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 669–674. University of Miami, 2011.
- [12] Ellis, D., Berenzweig, A., and Whitman, B. The “us-pop2002” Pop Music data set, 2003.
- [13] Y. Han, J. Kim, and K. Lee. Deep convolutional neural networks for predominant instrument recognition in polyphonic music. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(1):208–221, 2017.
- [14] D.E. Hast and S. Scott. *Music in Ireland: Experiencing Music, Expressing Culture*. Oxford University Press, Oxford, UK, 2004.
- [15] N. Keegan. The Parameters of Style in Irish Traditional Music. *Inbhear, Journal of Irish Music and Dance*, 1(1):63–96, 2010.
- [16] Y.E. Kim, D.S. Williamson, and S. Pilli. Towards Quantifying the “Album Effect” in Artist Identification. In *Proceedings of the 7th International Conference on Music Information Retrieval*, pages 393–394, Victoria, Canada, 2006.
- [17] G. Larsen. *The essential guide to Irish flute and tin whistle*. Mel Bay Publications, Pacific, Missouri, USA, 2003.
- [18] H. Lee, P. Pham, Y. Largman, and A.Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems*, pages 1096–1104, 2009.
- [19] T. Li, A.B. Chan, and A. Chun. Automatic musical pattern feature extraction using convolutional neural network. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2010.
- [20] T. Li and M. Ogihara. Music artist style identification by semi-supervised learning from both lyrics and content. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 364–367. ACM, 2004.
- [21] T. Lidy and A. Schindler. Parallel convolutional neural networks for music genre and mood classification. *MIREX*, 2016.
- [22] B. Logan. Mel Frequency Cepstral Coefficients for Music Modeling. In *Proceedings of the 1st International Symposium on Music Information Retrieval*, Plymouth, MA, USA, 2000.
- [23] M.I. Mandel and D. Ellis. Song-Level Features and Support Vector Machines for Music Classification. In *Proceedings of the 6th International Conference on Music Information Retrieval*, volume 2005, pages 594–599, London, UK, 2005.
- [24] L. E McCullough. Style in traditional Irish music. *Ethnomusicology*, 21(1):85–97, 1977.
- [25] T. Ó Canainn. *Traditional Music in Ireland*. Ossian, Cork, Ireland, 1978.
- [26] F. Pachet and J. Aucouturier. Improving timbre similarity: How high is the sky. *Journal of negative results in speech and audio sciences*, 1(1):1–13, 2004.
- [27] T. Park and T. Lee. Musical instrument sound classification with deep convolutional neural network using feature fusion approach. *CoRR*, abs/1512.07370, 2015.
- [28] D. Sussillo. Random walks: Training very deep non-linear feed-forward networks with smart initialization. *CoRR*, abs/1412.6558, 2014.
- [29] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*, 10(5):293–302, 2002.
- [30] F. Vallely. *The Companion to Irish Traditional Music*. Cork University Press, Cork, Ireland, 2011.
- [31] B. Whitman, G. Flake, and S. Lawrence. Artist detection in music with minnowmatch. In *IEEE workshop on Neural Networks for Signal Processing*, pages 559–568, Falmouth, Massachusetts, USA, 2001. IEEE.
- [32] G. Widholm, R. Linortner, W. Kausel, and M. Bertsch. Silver, gold, platinum-and the sound of the flute. In *Proc. Int. Symposium on Musical Acoustics, Perugia*, pages 277–280, 2001.
- [33] S. Williams. *Irish Traditional Music*. Routledge, Abingdon, Oxon, 2010.

		Predicted Player					
		1	2	3	4	5	6
Actual Player	1	0.9881	0	0	0	0	0
	2	0	1.0000	0	0	0	0
	3	0	0	1.0000	0	0.0063	0
	4	0	0	0	1.0000	0.0089	0
	5	0.0119	0	0	0	0.9848	0
	6	0	0	0	0	0	1.0000

Table 2.
Evaluation = 1.0
Confusion Matrix showing 6 individual players

		Predicted Player						
		1	2	3	4	5	6	Other
Actual Player	1	0.9600	0.0076	0	0.0172	0	0	0
	2	0.0066	0.9784	0	0	0	0	0.0083
	3	0.083	0	0.9792	0	0.0063	0	0.0337
	4	0	0	0	0.9828	0	0	0.0234
	5	0.251	0	0.0208	0	0.9875	0	0.0042
	6	0	0	0	0	0	1.0000	0.0125
	Other	0	0.0140	0	0	0.0063	0	0.9179

Table 3.
Evaluation 2
Confusion Matrix showing 6 individual players and 'other' class of 79