**Machine Annotation of   
Traditional Irish Music**

**PhD**

**Bryan Duggan MSc, BSc.**

*School of Computing*

*Dublin Institute of Technology*

*Kevin St. Dublin 8, Ireland.*

**Supervisors:  
Prof. Brendan O’ Shea  
Dr. Mikel Gainza**

**Prof. Padraic Cunningham**

**Dublin Institute of Technology, School of Computing**

**December 2008**

**Abstract**

The document contains a discussion of the characteristics of creativity and style in traditional Irish flute music and discusses related work on the problem of feature extraction from monophonic recordings of traditional music. Work on the modelling of musical creativity and the cognition of musical style in software is also presented.

**Keywords: Musical creativity, musical style modelling, traditional Irish music, transcription, signal processing, Machine learning**

**Table of Contents**

[1 Introduction 1](#_Toc203913804)

[1.1 Background 1](#_Toc203913805)

[1.2 Research aims 3](#_Toc203913806)

[1.3 Original Contribution 4](#_Toc203913807)

[1.4 Organisation 4](#_Toc203913808)

[2 Traditional Irish Music 6](#_Toc203913809)

[2.1 Tune Types 6](#_Toc203913810)

[2.2 Instruments 6](#_Toc203913811)

[2.3 Musical Creativity 6](#_Toc203913812)

[2.4 Style in Traditional Flute Music 9](#_Toc203913813)

[2.4.1 Ornamentation 9](#_Toc203913814)

[2.4.2 Breathing 12](#_Toc203913815)

[2.4.3 Regional Styles 13](#_Toc203913816)

[2.5 Conclusions 15](#_Toc203913817)

[3 Content Based Music Information Retrieval 17](#_Toc203913818)

[3.1 Searching symbolic representations 17](#_Toc203913819)

[3.2 Searching audio data 18](#_Toc203913820)

[3.3 Hybrid approaches 18](#_Toc203913821)

[3.4 Feature Extraction 20](#_Toc203913822)

[3.4.1 Pitch 20](#_Toc203913823)

[3.4.2 Note onset detection 20](#_Toc203913824)

[3.4.3 Loudness 24](#_Toc203913825)

[3.4.4 Chroma 24](#_Toc203913826)

[3.4.5 Timbre 24](#_Toc203913827)

[3.4.6 Mel-Filtered Cepstral Coefficients 24](#_Toc203913828)

[3.4.7 Wavelet analysis 24](#_Toc203913829)

[3.4.8 Spectral Centroid 24](#_Toc203913830)

[3.4.9 Ornamentation 24](#_Toc203913831)

[3.5 Melodic Similarity 24](#_Toc203913832)

[3.5.1 Parsons’ Code 24](#_Toc203913833)

[3.5.2 Intervals 25](#_Toc203913834)

[3.5.3 Earth Movers Distance 25](#_Toc203913835)

[3.5.4 Edit Distance 25](#_Toc203913836)

[3.6 Conclusions 26](#_Toc203913837)

[4 Machine Annotation of Traditional Tunes (MATT2) 28](#_Toc203913838)

[5 Machine Annotation of Traditional Sets (MATS) 42](#_Toc203913839)

[6 Conclusions 55](#_Toc203913840)

[6.1 Conclusions 55](#_Toc203913841)

[Bibliography 57](#_Toc203913842)

**Table of Figures**

Figure 1: Wooden flutes (Source: Author) 2

Figure 2: Characteristics required in a human listener for the cognition of musical style (Source: Author based on (Baroni 2006)) 8

Figure 3: An example of a *run* in ABC format (Source: Author) 12

Figure 4: Geographic origin of regional style (Source: Author based on (Keegan 1992)) 14

Figure 5: Waveform plots of a piano (top) and a wooden flute (bottom) playing the notes A to G (Source: Author) 21

Figure 6: the Onset Detection Function (ODF) for a musical phrase calculated using the Onset Detection using Comb Filters implemented by the author in Java (Source: Author) 22

Figure 7: The tune "Come West Along the Road" in the ABC format [11]. 29

Figure 8: High level diagram of the MATT2 tune annotation system 32

Figure 9: Signal and ODF plots of the first bar of the tune "The Boyne Hunt" 33

Figure 10: *Fuzzy histogram* of candidate note lengths from a 28 second phrase from the tune "The Hunters Purse" 35

Figure 11: Pseudocode for the fuzzy histogram quaver length calculator 36

Figure 12: Normalisation stages for the A part of the tune “Come West Along the Road” 37

Figure 13: Screenshot of MATT2 38

Figure 14: Top ten edit distances for a recording of the tune "The Golden Keyboard" 40

Figure 15: Waveform of the last phrase from the tune "Jim Coleman’s" and the first phrase from the tune "George Whites Favourite" played in a set 44

 Figure 16: High level diagram of the MATT2 tune annotator 46

Figure 17: Pseudocode for the MATS set annotation algorithm 50

Figure 18: Edit distance profiles for three tunes played in a set 51

Figure 19: Filtered version of first graph in Figure 4. The dynamic threshold and detected troughs are marked 51

**Table of Tables**

Table 1: Possible features that characterise creativity in traditional Irish flute playing (Source: Author) 13

1. Introduction

This document presents a summary of work carried out towards a PhD on the topic of modelling the cognition of musical creativity. The work specifically seeks to model the cognition of the creative interpretation of traditional Irish music on the concert flute. Related work proposes that musical performance involves tacit knowledge about interpretation that humans acquire by observation and imitation (DeManteras and Arcos 2002). This research examines the possibility of simulating this observation process using signal processing and machine learning in order to synthesise a novel approach to computationally modelling the cognition of the creative interpretation of traditional Irish music.

This work will have several practical applications in musician and regional style identification (multi-media indexing) in addition to applications in the creation of royalty free music and pedagogical applications in music teaching. Additionally the work will contribute to an understanding of the cognitive processes underlying musical creativity and the cognition thereof and it may lead to better modelling of the regional and cultural context of musical creativity (Csikszentmihalyi 1999). Although this work focuses on traditional Irish music as played on the concert flute, it is hoped that the techniques proposed can be generalised to other genres and instruments.

* 1. Background

Irish traditional music includes several musical forms. In the song tradition, both sean nós (“old style” singing in the Irish language) and singing in English exist. The baroque music of Turlough O'Carolan is also considered part of the tradition (Vallely 1999). This project however, is primarily concerned with traditional dance music, as played on the concert flute. The most common forms of dance music are *reels, double jigs* and *hornpipes*. Other tune types include *marches, set dances, polkas, mazurkas, slip jigs, single jigs and reels, flings, highlands, scottisches, barn dances, strathspeys* and *waltzes* (Larson 2003). These forms differ in time signature, tempo and structure. For example a reel is generally played at a lively tempo and is in 4/4 time (although played and transcribed as 8 quavers in a bar) while a waltz is generally played at slower pace and is in 3/4 time. The time signature, tempo and structure of a tune form are determined by the dance it accompanies. Most tunes consist of a common structure of two parts called either the first and second part or the A part and B part. Tunes are typically arranged into sets. A set consists of a number of tunes (commonly two or three) played sequentially. Each tune in a set is usually repeated two or three times (Vallely 1999).

Instruments used to play traditional dance music include the tin whistle, fiddle (violin), uilleann (elbow) pipes, accordion, concertina, harp and the banjo (Wallis and Wilson 2001). The flute came into common use in traditional music in the 19th Century. The “Irish flute” is also known as the concert flute (because it is in concert pitch), the timber flute (because it is made from wood), the simple system flute or the fheadóg mhór (big whistle). It has six holes tuned such that the lowest playable pitch (all holes closed) is the D above middle C, and the instrument will play a D scale (D, E, F#, G, A, B, C#) as the holes are uncovered sequentially to shorten the resonant length of the bore. The basic flute is often augmented with the addition of up to eight keys (typically made from silver, mounted on wooden blocks) used to play pitches which are impossible to produce on the basic flute. Figure 1 depicts a 6 keyed wooden flute made from African black wood by Eamonn Cotter, an unkeyed made from African black wood flute made by Eamonn Cotter and an unkeyed bamboo flute made by Patrick Olwell in the key of F.



Figure : Wooden flutes (Source: Author)

Wooden flutes from the 19th Century were originally designed to play classical music, but with the invention of the Boehm system flute in 1832, wooden flutes became unpopular amongst classical musicians and thus came to be acquired by traditional musicians. Since the 1970’s, there has been a renaissance in wooden flute making and now many musicians play modern wooden flutes based on the 19th Century designs (Vallely 1999).

Music is a creative art form and “individual expression” is a defining component of traditional Irish music (Breathnach 1977). Creativity in traditional music takes three forms:

1. The composition of new tunes.

2. The arrangement of tunes into sets.

3. The individual creativity of a musician in interpreting a tune.

This work concentrates on individual, interpretive creativity. When a traditional musician plays a tune, it is rarely played exactly as transcribed, though unlike with jazz for example, traditional musicians never deviate from the structure or framework of the tune. In fact an experienced musician rarely plays the same tune twice, identically. Interestingly, there is no scope in traditional dance music for rubatto (except for micro-tempo artefacts). Instead, a musician will employ the subtleties of ornamentation and variation to interpret the tune (Larson 2003).

Ornamentation plays a key role in the individual interpretation of traditional Irish music (Canainn 1978). Ornamentation has a different meaning in Irish traditional music than its definition in classical music. In classical music, the expression is achieved by adding notes to the melody. By contrast, with the exception of the slide effects, Irish traditional music ornamentation is played on the beat, and alters the onset of the notes in a manner in which, it is argued, only one note will be heard (as opposed to two notes as in classical music) (Larson 2003). The usage of ornamentation is highly personal and large variations exist in the employment of ornamentation from region to region, instrument to instrument and from musician to musician.

* 1. Research aims

The principal aim of this research is to synthesise an approach to modelling the cognition of musical creativity in traditional Irish flute music. The model will be developed as software that, based on human cognition of musical style can identify and classify creativity and style in digital recordings of traditional musicians.

To achieve this aim, it will first be necessary to understand how humans achieve this task. A literature review will be carried out on the combined subjects of creativity and style in traditional flute playing and the modelling of musical creativity in software. This will be supported by an experiment which will seek to establish definitively the feature set used by domain experts to classify traditional flute styles.

As will be demonstrated, the extraction of high level features from digital recordings of traditional music is a significant and ongoing research challenge. It will be necessary therefore to establish the state of the art in this field and integrate and expand this work into a combined approach to modelling the cognition of musical creativity. Machine learning algorithms will be used to build a classifier from the feature set generated by the analysis of recordings and the aim is to use standard techniques such as cross fold validation to test the model.

* 1. Original Contribution

The attempt to automate the cognition of creativity and musical style in Irish traditional flute music represents a novel contribution in the field, since modelling this type of music cognition has never been attempted previously. In addition, the main specific contributions to knowledge are listed as follows:

1. The identification of a set of creative features that distinguish the interpretation of a piece of traditional music by one musician from another.
2. The development of new algorithms for high level feature extraction from recordings of traditional music.
3. The synthesis of new similarity metrics for automating the comparison of digital recordings of traditional flute music.
4. The synthesis of a novel approach to the automation of musician, style and region of origin identification from digital music recordings.
5. The synthesis of novel frameworks for musical style modelling that uses signal processing techniques to infer high level stylistic features that can be modelled by a machine learning system.
6. The validation of all frameworks proposed.
   1. Organisation

The remaining sections of this document are organised as follows:

Section 2 discusses creativity and style in traditional Irish flute music. Firstly, musical creativity and style are defined and related. Requirements for the cognition of style in a human listener are proposed. This section also summarises the characteristics of style in traditional flute playing, discussing ornamentation, breathing and regional styles.

Section 3 describes related work in two areas. In the field of feature extraction, signal processing techniques, using filters and Fourier analysis to determine note onset and offset times and pitches of notes in traditional Irish music are described. This section continues by reporting on a selection of systems that use machine learning to either model musical creativity or the cognition of musical style in software. Most of the systems discussed operate in the domain of classical music, though one of the systems discussed uses a corpus of traditional reels to learn from.

Section 4 discusses the proposed approach. A high level system diagram is presented illustrating how signal processing will be used to extract features from digital recordings of flute music in order to build a training corpus for a machine learning system. Four high level problems are identified in this section as needing to be addressed for the overall project aim to be achieved.

Section 5 presents the work carried out to date. Three musicians are identified who have agreed to assist in this project. The MATT (Machine Learning Articulation of Traditional Tunes) project developed by the author is described and discussed in this section as is current work on the problem of feature extraction.

Section 6 presents the work plan for the remainder of the PhD and identifies six high level tasks that will be carried out.

1. Traditional Irish Music

(Williamson, Thompson et al. 2006) identify reasons why authors have had difficulty characterising creativity. They suggest that it has been impossible to offer an unambiguous and broadly agreed on definition. Further they propose that creativity is difficult to isolate empirically and finally they suggest that creativity has an entrenched mythology especially in the arts world where it is construed as a mysterious, unknowable process. In this section, the problem of defining style in traditional flute playing is divided into two sub-problems. Firstly the concept that style is related musical creativity is proposed. Secondly, this section summarises approaches to the problem of what characterises style in traditional flute playing.

* 1. Tune Types
  2. Instruments
  3. Collections
  4. Musical Creativity

(Götz 1981) relates creativity to “making” and defines creativity as “the process or activity of deliberately concretising insight”. (Boden 1996) is extensively cited by authors seeking to understand creativity. Boden distinguishes two types of creativity. Psychological creativity (P-creativity) occurs when an individual has an idea which is novel to that individual, regardless of how many other individuals have had that same idea. Historical creativity (H-creativity) defines ideas that are novel not only to an individual, but also novel in the history of human endeavour. P-creativity is therefore judged by an individual. H-creativity is judged by society at large. The concept of two levels of creativity is also proposed by (Gardner, 1993b), who distinguished between “little c” and “big C” creativity.

There are examples in traditional music of both P-creativity and H-creativity as defined by (Boden 1996). Individual expression (P-creativity) is in fact a defining component of traditional Irish music (Breathnach 1977). When a traditional musician plays a tune, it is rarely played exactly as transcribed, though unlike with jazz for example, traditional musicians never deviate from the structure or framework of the tune. In fact, experienced musicians rarely play the same tune twice, identically. In the introduction to the revised edition of O’ Neill’s Music Of Ireland (originally published in 1906), Krassen describes a typical scenario:

“*It seems that on this particular occasion Touhey wanted to learn a tune from McFadden. He had McFadden play it for him several times and then tried his own hand at it. Of course McFadden had to play it again, pointing out several "errors." This happened a number of times until Touhey finally gave up, for McFadden was playing the tune a little differently each time through!*”

- (Krassen 1975)

A traditional musician will usually employ subtle variations, ornamentation, timbre and phrasing to interpret a tune (Larson 2003).

H-creativity by definition, more rarely occurs in traditional music. Some examples might include the introduction of the concert flute in the nineteenth century, the development of the ceili band form in the 1920’s, the renaissance of traditional music led by Sean O’ Riada and Ceolteori Cuailann in the 1960’s and the introduction of the Bouzuki in the 1970’s (Wallis and Wilson 2001).

The cognition of individual creativity implies that an individual musician demonstrates a style that can be recognised. (Meyer 1989) defines musical style as:

“*a replication of patterning…that results from a series of choices made within some set of constraints”*.

- (Meyer 1989)

(Keegan 1992) again associates the concept of style with creativity and claims that the technique and creativity of an individual and their musical style are one and the same thing.

(Baroni 2006) suggests that a listener can have different approaches to music which influences their perception of style. A listener’s approach can be:

*“a mere abandon to the flux of sounds where music is lived as an emotional stimulus and a source of immediate pleasure”*.

- (Baroni 2006)

He suggests that in this context a listener has little appreciation of the style of the musician. He continues by proposing that a listener must have an *objective approach*, *a precise knowledge* of the cultural conditions where the music was produced and must have *competence* to distinguish one style of music from another. He suggests that examples are categorised by comparing them to *prototype models* that represent its fundamental characteristics.



Figure : Characteristics required in a human listener for the cognition of musical style (Source: Author based on (Baroni 2006))

(Baroni 2006) describes an experiment carried out to establish the features used by a group of both experts and amateurs in a musical domain to categorise a piece of music. A group of 13 subjects listened to a recording of a fragment of a little known piece of music by the composer Donizetti. The subject group contained musicologists, professional and amateur musicians. Each subject was provided with a tape recorder to record the cognitive paths followed in order to identify the composer. The experiment demonstrated that those subjects who possessed “prototype models” or “stored memories” were able to identify the century and genre, form and instruments in the piece of music, in other words to classify the style. The experiment also demonstrated that those subjects who possessed a “lexicon” of music terminology were better able to classify the piece. The author concludes by explaining that the subjects used:

*“prototype, conceived as a hierarchical organisation of memorised listening experiences, orientated by historical knowledge”*

*-* (Baroni 2006)

to classify the music.

* 1. Style in Traditional Flute Music

There are a number of authoritive sources that describe characteristics that can define an individual musician’s flute style. These include Valley’s, “Timber: The Flute Tutor”, and his PhD thesis, “Flute Routes to 21st Century Ireland” (Vallely 2004), Larson’s “The Essential Guide to Irish Flute and Tin Whistle”, McCormack’s, “Fliúit: Irish Flute Tutorial”, Keegan’s MPhil thesis “Words of Traditional Flute Style” (Keegan 1992). In addition there is Casey’s “Traditional Irish Flute Music from East Galway A Regional study and Documentary Field Collection”. Additionally Tansey’s “The Bardic Apostles of Inishfree” (Tansey 1999), a profile of Sligo musicians contains references to ornaments not described in any of the other literature, (*bark*, *backstitch, run* and *pop*). In personal interviews (Tansey 2006) he has elaborated on the meaning of these terms. Although there are some disagreements in definitions of certain features, the literature generally agrees that flute style can be characterised by features that include use of ornamentation, phrasing (where a musician takes a breath), use of variation, staccato or legato playing (with throating/tounging attacks), the timbre a musician achieves with an instrument, tempo, choice of tune and choice of tune type.

* + 1. Ornamentation

(Larson 2003) defines ornamentation as:

“…*ways of altering or embellishing small pieces or cells of a melody that are between one and three eight-note beats long. These alterations and embellishments are created mainly through the use of special fingered articulations*.”

- (Larson 2003)

Fingered articulations are a defining characteristic of traditional Irish music. The sound of most articulations is very brief. Although generated by inserting additional notes, (Larson 2003) argues that the notes are played at such speed that they are not perceived as having a discernible pitch or duration. There are differing opinions as to the origins of ornamentation in traditional Irish music. (Larson 2003) suggests that ornamentation is derived from the playing of the *píob mór*, a mouth blown bagpipe that predated the development of the modern uilleann pipes. The  *píob mór* had no capacity for momentary interruptions to the flow of air and thus melodies were played as unbroken streams of sound. In order to generate a perceived stop between two notes of the same pitch, a musician would play a third note momentarily between the two notes.

(Tansey 1999) argues that ornamentation developed as an attempt to mimic the sounds of nature. He compares for example the sound of a *cran* to that of a sheep’s “baa” and postulates that the ornament was developed by shepherd’s who played wooden flutes while tending sheep:

“*I put it to you therefore that it had to come from the throats of birds, the wild animals, the ancient chants of our forefathers, the hum of the bees and the mighty rhythms of the galloping hooves of wild horses all moulded together…*”

- (Tansey 1999)

The main components of wooden flute ornamentation are now identified:

A *cut* is defined as an articulation used to separate two notes. A cut is articulated by playing a middle note momentarily at a higher pitch than the second note. The overall length of the two notes does not change when cutting and so the length of the second note must be shortened very slightly to accommodate the cut.

A *tap* (referred to in some sources as a *strike* or a *bounce*) is an articulation also used to separate two notes. A tap is articulated by playing a middle note momentarily at a lower pitch than the second note.

A *long roll* is articulation used to separate three notes. The second note in the sequence is cut and the third note is tapped. Again, the overall length of the three notes does not change. A *short roll* is similar to a long roll, but the first note in the sequence of three is dropped.

Concert flutes are usually pitched in D. As there is no note lower than a low D on the instrument, a tap on the low D is not possible. Instead, to execute a “roll” type ornament on a low D, a musician will play a *cran*. In order to play a cran, the musician replaces the tap with a second cut. The second cut uses a different note, usually higher than that of the first cut. This creates a “bubbling” sound typical of the playing of Matt Molloy. Not all musicians use crans, for example Catherine McEvoy does not play crans at all. Although (Larson 2003) suggests that crans can be done on any note, most other sources suggest that crans are only played on the low and middle D and E. They can be played long or short as with rolls.

With all of the above articulations, the actual pitch of the “extra” notes may vary depending on which finger the musician feels most comfortable lifting at speed (Keegan 1992). Using different fingers to perform the ornamentation also gives the ornament a specific character which can be part of a musician’s unique sound. An interesting example of this can be found Seamus Tansey’s 1975 recording “The King of the Concert Flute” (Tansey 1975).

(Larson 2003) suggests that trills are not common in Irish flute music, however an analysis of the corpus described in section **Error! Reference source not found.** finds this not to be the case. A trill is defined as a rapid alteration of the principal note and the note above it. A trill may begin on either the principal note or on the higher ornamental note. Trills are usually played for short durations in traditional music, with longer duration trills being considered too much of an allusion to classical music.

A *tight triplet* alsocalled a *treble* in (Tansey 1999) is a stepwise rising or falling sequence of three notes played in quick succession in the rhythm of two notes. A specific type of tight triplet mentioned in (Tansey 1999) is a *back stich* which he describes as a treble using the notes BCD. A *run* as described by (Tansey 2006) is a descending sequence of two tight triplets as illustrated in Figure 3. In the note sequence, the first four notes are played without the use of a run while the second sequence of six notes are two tight triplets, in other words a *run* on the four note sequence.

K:D

M:Reel

=cABG (3=cBA (3BAG

Figure : An example of a *run* in ABC format (Source: Author)

Switching between octaves on a wooden flute is achieved using a technique known as *overblowing*. (Hamilton 1990) describes how overblowing can be used as a technique to add variation to a performance by overblowing a phrase meant to be played in the lower octave of the instrument.

Overblowing is also used as a technique in the sounding of a *hard D*. A *hard D* is achieved on a wooden flute by overblowing the D in the lower register to the extent that the note is perceived as a group of harmonics of D that can be impossible to distinguish (Keegan 1992).

* + 1. Breathing

Phrasing in traditional flute music is easily identified as the timings in a performance of a tune where a musician takes a breath.

|  |  |  |
| --- | --- | --- |
| Ornamentation | Single-note | Cut |
| Tap |
| Multi-note | Roll |
| Cran |
| Triplet |
| Run |
| Breathing | Phrasing |  |
| Throating (attacks) |  |
| Overblowing |  |
| Timbre |  |
| Variation |  |  |
| Repertoire | Reels |  |
| Jigs |  |
| Hornpipes |  |
| Polkas |  |
| Slides |  |
| Scotisches |  |
| Strathspeys |  |
| Mazurkas |  |
| Tempo |  |  |

Table : Possible features that characterise creativity in traditional Irish flute playing (Source: Author)

Traditional music scores are not annotated with breath marks and it is up to an individual musician to decide where a breath should be taken. Taking a breath usually means leaving out a note or several notes from the score in a performance. Phrasing is therefore more obvious in music played on the flute than on any other traditional instrument (Keegan 1992). (Keegan 1992) in his interviews establishes that phrasing (and in particular the length of phrases) is a strong indicator of a particular regional and individual style.

In traditional Irish flute playing, *tounging* as used as a note attack by classical flute players is rarely used. Instead a technique called *throathing* is often used (the stop is produced by the throat rather than by the tongue) (Hamilton 1990). This can often result in the note following the attack to be overblown, (sometimes one of the harmonics of the fundamental rather than the fundamental itself is perceived).

On the flute, the timbre achieved by a musician can vary widely between a broad/breathy sound and a sharp/clear sound and naturally, volume also can characterise a style.

Table 1 summarises the possible features elaborated upon in this section.

* + 1. Regional Styles

(Canainn 1978) describes regional style as the common features which distinguish the majority of performances by musicians from a particular area. Until the 1940’s there existed distinct regional styles of flute playing attributed mainly to the isolation of rural communities prior to the advent of mass communication. Similarly the country as a whole was largely preserved from the influence of other cultures due to its geographic position and the isolationist economic policies of the early Irish Free State (Keegan 1992). (Keegan 1992) describes his work in understanding the cognition of regional styles of Irish flute music by conducting a series of interviews with prominent musicians. He reports that four regional styles were identified by his subjects, though his work suggests that the characteristics that distinguished these styles varied somewhat. The regional styles identified in his work are: The West Clare style, the Ballinakill/East Galway style, the Fermanagh/Northern style and the Sligo/Roscommon style. Figure 4 shows a map of Ireland with the locations of the four regions identified by (Keegan 1992).

The West Clare and Ballinakill/East Galway styles he describes as demonstrating much use of ornamentation and accidentals, with the melody played at a relatively slow pace. These styles differ in repertoire and use of breath articulation, with The West Clare style being characterised by the use of throathing to emphasise rhythm. The Ballinakill/East Galway style developed from the playing of the musicians in one of the first ceili bands (The Ballinakill Traditional Players). (Keegan 1992) suggests that the Ballinakill/East Galway sound is more legato, with an emphasis on melody rather than rhythm. This is evident in the repertoire played by musicians in that style, which contains tunes with several parts. The suggests that in the past a substantial group of East Galway musicians have adopted the Boeme system flute or other fully keyed instruments, which are more suitable for the repertoire which involve tunes in unusual keys and with accidentals.



Figure : Geographic origin of regional style (Source: Author based on (Keegan 1992))

The Fermanagh/Northern style he describes as being sparsely-ornamented, but with heavy stress on breath articulation techniques. He states that there exists two styles of phrasing. In some examples, there is an emphasis on natural-phrasing (regular two-bar phrases), while other musicians demonstrate short irregular phrasing, characteristic of the music of North Leitrim (and hence similar to the Sligo-Roscommon style).

There is a strong concentration of flute players in the Leitrim/Sligo/Roscommon area which (Tansey 2006) attributes to the prevalence of coal mining in the region. He argues that the flute was considered good for the development and health of the lungs of coal miners, constantly exposed to high levels of coal dust in their profession. Although (Keegan 1992)’s subjects reported contradictory opinions on many aspects of the Sligo/Roscommon style, they agreed that the style is very rhythmical because of the use of breath articulation and emphasis. They also suggest that the overuse of ornamentation is not characteristic of many musicians of the Sligo/Roscommon style (though he points out several notable exceptions).

* 1. Conclusions

From this section, it is possible to conclude that the concept of style and creativity are strongly related. A musician’s style can be described as the creative choices made in the performance of a piece of music. From the perspective of traditional flute playing, these choices include the use of ornamentation, phrasing, use of variation, staccato or legato playing with correspondingly different attacks, the timbre a musician achieves with an instrument, tempo, choice of tune and choice of tune type. It is clear also that flute style has a regional context and how a musician interprets tune is often strongly influenced by the cultural context the musician has emerged from. The cognition of musical style requires competence, objectivity and knowledge of this cultural context in listeners. Finally, it is proposed that the cognition of style involves the listener comparing an unclassified piece with prototype models.

Support for queries on traditional instruments or lilted queries

Ornamentation

Variation including scattering

Phrasing (breathing)

Sets

Legato playing (hard to detect onsets)

Repetition & Structure (no reason for short queries or incipit's)

Tempo variation

Any phrase (not just the incipit)

Transcription errors (false positives, false negatives, pitches)

The importance of the problem in the domain of cultural heritage preservation

Transposition?

1. Content Based Music Information Retrieval

The approaches proposed in this work form part of a content based Music Information Retrieval (MIR) system for traditional Irish music. Music Information Retrieval can be defined as “the task of extracting from a large quantity of musical data, the portions of that data with respect to which some musicological statement is true” (Kassler 1966). The term Music Information Retrieval is first mentioned in the literature in (Kassler 1966). In this work the author presents MIR, an assembly like language for formulating musical queries and navigating scores. He suggests that MIR could form part of a “library of the future” although he recognises the limitations of the language proposed.

More recently (Typke, Wiering & Veltkamp 2005; Typke 2007) suggest that there are three main classifications of MIR systems: those for searching symbolic representations of music, those for searching audio data and systems that combine both approaches by first converting audio data to a symbolic representation and then searching for a match in a corpus of symbolically notated music. (Downie 2003) proposes analytic/production systems and locating MIR systems a classification analogous to the first two classifications. This section presents related work in each of the three classifications of system and concludes with an analysis of the suitability of existing approaches to perform MIR for traditional Irish music.

* 1. Searching symbolic representations

Symbolic MIR has its roots in dictionaries of musical themes such as Barlow [1]. Monophonic music can be represented as a one-dimensional string of characters, where each character represents a musical note. String can be made up of characters representing pitches, pitch intervals or melody contours. In systems that use this format, standard string matching algorithms such as Knuth-Morris-Pratt, Boyer-Moore, Levenstein (Edit) Distance, longest common sub-sequence or regular expression searching have been applied (Navarro & Raffinot 2002).

*GuidoMIR* is a symbolic MIR system that has a native corpus of melodies in the Guido/XML music notation language. The authors claim that using a symbolic musical score language such as Guido/XML has a number of advantages over MIDI, a format designed for playback. They cite the ability to store meta-data with the melody as the main advantage, but list several others. They also do not use any form of database engine and instead their system in built entirely in Perl and uses a database of flat files. Although their corpus is text based, the authors use a probabilistic matching algorithm based on first order Markov chains to match queries to corpus strings. Their system supports queries based on both pitch and rhythm. (Hoos 2001).

*TunePal* is an MIR system whose main advantage is that it runs on a mobile device such as PDA or smartphone and so can be used in traditional music sessions and workshops. TunePal has a corpus of approximately five thousand traditional Irish dance melodies in ABC format. The system supports text queries on melodies or any of the meta-data such as tune name, type or composer. For melodic queries, the system requires a knowledge of the ABC language. It has an elementary query normalisation algorithm that normalises text queries into the same register and removes ornamentation from corpus strings, but otherwise it requires to exactly match strings from the corpus .

*Orpheus*

*Themefinder*

* 1. Searching audio data

*Shazam*

* 1. Hybrid approaches

*MELDEX* has a pitch tracking interface that allows users to sing queries. The system depends on the user separating each note by singing *da* or *ta*. The articulation of the consonant is used to detect the onset of each note. As queries were generated by humans, they naturally contained errors. has classified the errors into four types: Expansion, Compression, Repetition, and Omission.

MELDEX has a database of approximately 10,000 folk songs, compiled from the Essen collection (ref). The system uses the approximate string matching methodology of . This methodology was designed explicitly for the musicological analysis of melodic strings. Melody contour searches use interval direction method (section x.x). Matching melodies are ranked based on the degree of similarity between query and the items returned. Initially, MELDEX supported queries based on incipit’s , however subsequent improvements facilitated the matching of queries where the match occurs not only in the incipit, but also anywhere within a melody . Reported performance of the system is quite poor, with simple, exact match searches, taking an average of 500 ms to perform and 20 note approximate search pattern, requiring approximately 21 seconds.

*Musipedia* (previously known as *Tuneserver*) is a web-based MIR system that supports queries entered by whistling, playing on a virtual piano keyboard, tapping the rhythm on the computer keyboard, or entering the melodic contour . For whistled input, the audio is first sampled and a Fast Fourier Transform is then used estimate pitch. Onsets are noted using a combination of *silence windows* and pitch changes between consecutive frames of audio. The audio is then converted to Parsons’ code (section X.x) and a melodic contour search calculates the weighted edit distances between the query and strings from the corpus. Results are ranked in order of ascending distance from the query. The authors report a success rate of approximately 80% for queries with an average of 16 notes, where the correct melody was within the top 40 matches. The correct melody was returned as the closest match in 44% of queries. The authors ascribe mistakes to transcription errors and queries that were too short to discriminate similar representations of different melodies. The front end to Musipedia is also known as Melodyhound. Interestingly, although Musipedia contains traditional Irish dance tunes as part of its corpus, it does not generate positive results when queries are presented in the form of melodies played on the tin-whistle or wooden flute.

A later implementation of Musipedia supports a pitch and onset time-based search by representing the query into a weighted point set and calculating the Earth Mover's Distance (Section X.x) for each query point set and pre-computed point sets representing segments of melodies from the database. The "query by tapping" method that only takes the rhythm into account uses the same algorithm as the pitch and onset time method, but assumes all pitches to be the same. The system accelerates searches using an indexing technique based on vantage objects .

*Fast melody Finder* (FMF) is a web based music information retrieval prototype whose key feature is that it indexes the corpus according to a scheme known as FAI (Frequently Accessed Index) . The principal behind FAI is that a piece of music is often identifiable from a few specific melody segments of the overall melody. In FAI, segments are automatically induced from previous user queries. Each entry in the FAI structure has four variables: Access Count, Age, Repetition and Size. The authors propose an index maintenance system that, for example, supports merging of similar indexes.

Their prototype system has a corpus of 12000 MIDI files, which they pre-process to extract meta data in XML format such as time and key signature. Melodies are represented as pitch (U, D, S) and time contours (L, S, S) (Section X.x). Queries can be input by humming or by drawing the melody on a graphical representation of a 5 line stave. The system presumably incorporates a transcription subsystem, but this is not discussed in the paper. Matching is achieved using the Boyer Moore algorithm initially to search for an exact match and if an exact match is not found the system falls back to calculating the edit distance using dynamic programming. Index entries are searched in order of access count. The authors present results which indicate that queries using both pitch and time contours

(Blum

et al., 1999). Foote (1997) has designed a music indexing

system based on histograms of MFCC features derived

from a discriminatively trained vector quantizer.

Cornells Query By Humming

Super MBox

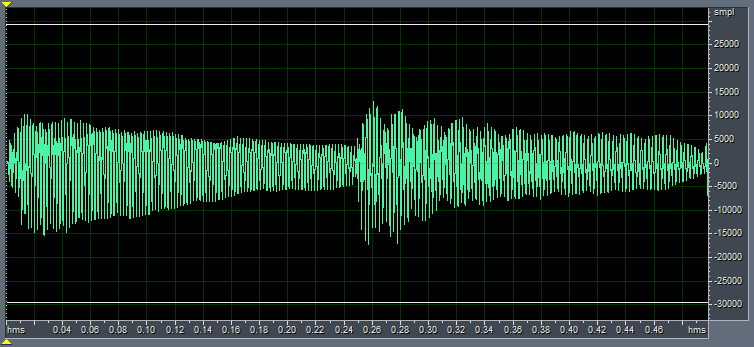
SoundCompas

* 1. Feature Extraction

Introduction blagh blagh blach

* + 1. Pitch
    2. Note onset detection

To address the first problem, the authors propose an approach they refer to as Onset Detection using Comb Filters (ODCF). ODFC discovers harmonic characteristics of the input signal and is therefore more tolerant to energy changes in an input signal and is also better at detecting onsets in legato playing, where there is no significant change in energy at the onset of a new note.



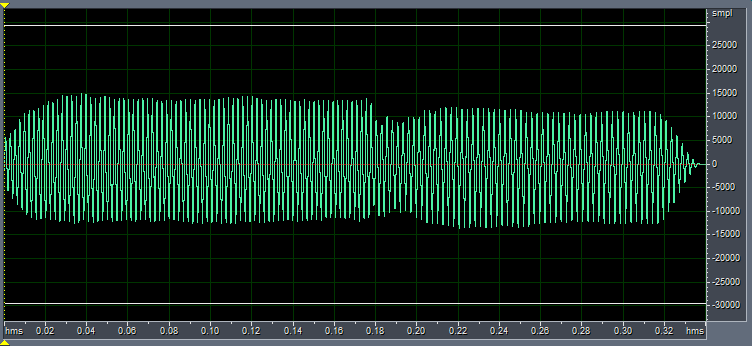


Figure : Waveform plots of a piano (top) and a wooden flute (bottom) playing the notes A to G (Source: Author)

Figure 5 compares waveform plots of a wooden flute playing the notes A to G legato with waveform plots of the same notes being played on a piano. As can be seen from this figure, there is a significant and detectible energy change in the plot from the piano between the offset of the first note and the onset of the second note, whereas with the notes played legato on the wooden flute there is a less detectable energy change from one note to the next.

To generate the Onset Detection Function (ODF), the input signal is first sampled at 44100Khz. The input signal is then segmented into overlapping frames of 2048 samples (approximately 46 milliseconds). Each frame overlaps with the previous frame by 75%. Each frame is then passed through a bank of twelve FIR comb filters.

A FIR comb filter works by summing the input signal with a delayed version of the same input signal. The delay of the filter is calculated as being 1 / frequency being filtered (the length in time of a single period of a waveform at the frequency). This has the effect of amplifying the frequency (or a harmonic thereof) in the input signal that matches the frequency being filtered. Thus, the energy of the input signal is doubled only if the peaks of the signal coincide with the peaks of the FIR comb filter. This will only occur for a given delay and its integer multiples.

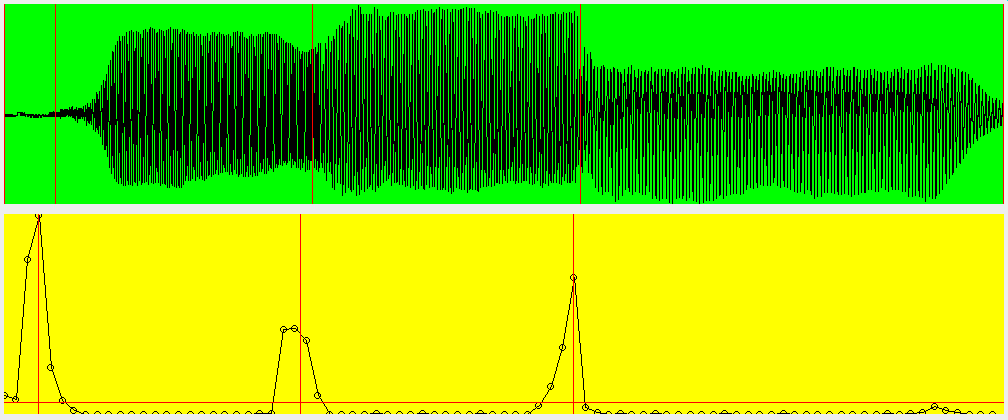


Figure : the Onset Detection Function (ODF) for a musical phrase calculated using the Onset Detection using Comb Filters implemented by the author in Java (Source: Author)

Twelve filters with different delays are used corresponding to the twelve semitones in the key of D3. For each frame of audio examined, the outputs of the audio passed through each of the twelve filters are calculated. A value for the ODF is then calculated as being the sum of the difference between the outputs of each of the twelve filters in successive frames, squared. In the case where the input signal changes from one note to another, this results in a peek in the ODF graph.

Using statistical techniques (average and standard deviation), a threshold is then calculated above which peeks in the ODF are recognised as being candidate note onsets. Figure 6 shows the ODF calculated in this way (using the system developed in Java and described in section **Error! Reference source not found.**) with an input signal of a wooden flute playing the notes D, E and F legato.

As illustrated, the onsets detected correspond to the onsets of each new note. The authors also propose that the ODF is filtered using a low pass filter to smooth the ODF, removing spurious onsets caused by noise in an onset.

(Dixon 2004) describes BeatRoot and the Performance Worm, two systems that the author claims, extract expressive features from a digital recording of a piece of music played by a human musician. They propose that although expression is contained in the physical features of the audio signal, such as amplitudes, frequencies and onset times, it is better understood when viewed from a higher level of abstraction, that is, in terms of musical constructs such as tempo, rhythm, pitch and timbre. These features are not directly measurable and the signal must be analysed to extract them.

BeatRoot models the perception of beats in a piece of music. BeatRoot first analyses the input signal to extract note onsets. Their first attempt to extract note onsets used a time domain algorithm that looked at the energy changes in successive frames. The authors claim that this approach worked well for percussive instruments such as the piano, but admit that the algorithm often detected false onsets and also failed to detect onsets for simultaneously sounding notes.

Their second attempt improves accuracy by separating the signal into frequency bands and looking for onsets in each band. (Gainza, Coyle et al. 2005) algorithm discussed earlier, which uses time domain comb filters seems more promising for detecting onsets in traditional music played legato on wind blown instruments as it is less sensitive to amplitude modulation in the signal. The system then uses an array of agents initialised with a tempo hypothesis. The agent then predicts further beats and is evaluated according to how well the predicted and actual beat times correspond. The system was evaluated against a corpus of Mozart sonatas and popular music and the authors claim a success rate of 90%.

* + 1. Loudness
    2. Chroma
    3. Timbre
    4. Mel-Filtered Cepstral Coefficients
    5. Wavelet analysis
    6. Spectral Centroid
    7. Ornamentation

To detect ornamentation, the algorithm use heuristics derived from standard descriptions of traditional ornamentation (summarised in section 2.5.1 of this document). For example, to transcribe a cut on the note G, the algorithm looks for two consecutive G notes, separated by a momentary note at a higher pitch. The authors report a 60% success rate with single note ornament and a 40% success rate at detecting multi-note ornaments.

* 1. Melodic Similarity
     1. Parsons’ Code

Parsons showed that a simple encoding of tunes that ignores most of the

information in the musical signal can still provide enough information for distinguishing between a large number of tunes. The Parsons code reflects only

the directions of melodies. Each pair of consecutive notes is coded as “U” (“up”)

if the second note is higher than the first note, “R” (“repeat”) if the pitches are

equal, and “D” (“down”) otherwise. Rhythm is completely ignored. Thus, the

first theme from the last movement of Beethoven’s 9th symphony (“Ode to Joy”)

would be coded RUURDDDDRUURDR.1 Note that the first note of any tune is

used only as a reference point and does not show up explicitly in the Parsons

code at all.

Lemstorm et al. (1998) used the suffix-tree as the index

and presented a coding scheme of music that is

invariant under different keys and tempos, and investigates

the application of two approximate matching algorithms

to retrieve music.

* + 1. Intervals

n-grams

* + 1. Earth Movers Distance
    2. Edit Distance

Edit distance, also known as *Levenshtein distance* or *evolutionary distance* (Levenshtein 1966; Navarro & Raffinot 2002) is a concept from information retrieval and it describes the number of edits (insertions, deletions and substitutions) that have to be made in order to change one string to another. It is the most common measure to expose the similarity between 2 strings.

The edit distance *ed(x, y)* between strings *x=x1 ... xm* and *y=y1 ... yn*, where *x, y* is the minimum cost of a sequence of editing steps required to convert *x* into *y.*  is the alphabet of possible characters and  is the set of all possible sequences of *ch* . Edit distance can be calculated using dynamic programming . To compute the edit distance *ed(x,y)* a matrix *M1...m+1,1...n+1* is constructed where *Mi,j* is the minimum number of edit operations needed to match *x1...i* to *y1...j*. Each matrix element *Mi,j*  is calculated as per (1). The minimum edit distance between *x* and *y* is given by the matrix entry at position *Mm+1,n+1*.

|  |  |  |
| --- | --- | --- |
|  | if *xi=yi*  else | (1) |

The algorithm can be adapted to find the lowest edit distances for *x* in substrings of *y*. This is achieved by setting *M1,j* = 0 for all *j* .*n+1.* In contrast to the edit distance algorithm described above, the last row *Mm+1,j* is then used to give a *sliding window* edit distance for *x* in substrings of *y* (Navarro & Raffinot 2002)*.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | D | G | G | G | D | G | B | D | E | F | G | A | B |
|  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| D | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 2 | 2 | 2 | 1 |
| E | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 1 | 0 | 1 | 2 | 3 | 2 |
| E | 4 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 2 | 1 | 1 | 2 | 3 | 3 |

Table : Edit distance for the string BDEE in DGGGDGBDEFGAB. This string represents the first 13 notes from the tune "Jim Coleman's" in normalised ABC format (section X.X)

An example of this variation on the edit distance applied to search for the pattern “BDEE” in “DGGGDGBDEFGAB” is given in Table 2. The minimum edit distance positions are highlighted. Variations on the edit distance algorithm have been applied in domains such as DNA analysis and automated spell checking and are commonly used in MIR systems (Birmingham et al. n.d.; Lemstrom & Perttu 2000; Rho & Hwang 2004; McPherson & Bainbridge 2001; Prechelt & Typke 2001).

Hidden Markov Models

Longest Common Subsequence see (Rho & Hwang 2004)

String matching algorithms

Boyer–Moore algorithm: This is based on the idea

that more information can be obtained by matching

the pattern from the right than from the left and shows

very good performance. It scans the pattern characters

for a match starting from the last character in the string.

During the search, the pattern characters are scanned

for a match starting with the last character in the

pattern.

* 1. Conclusions

MIDI – No support for meta data or structure (Hoos 2001)

Contours/intervals – Too many false positives (Schlichte 1990) (Adams, Bartsch, & Wakefield 2003)

current UDR string cannot describe sudden

pitch transitions

EMD – (No ornamentation compensation, cant use for segmentation - TYPKE 2007)

Musipedia, Melody hound etc (No support for query by playing)

None of the systems support segue melodies

None of the systems address ornamentation

1. Machine Annotation of Traditional Tunes (MATT2)

Repertoire in Irish traditional music is primarily acquired orally. Musicians playing Irish music learn tunes by hearing the tune played by fellow musicians in sessions, classes, workshops and from commercial recordings . It is common at workshops such as those held as part of the Willie Clancy Summer School for students to use electronic devices to record their classes. Increasingly students use digital audio field recorders such as the M-Audio Micro Track II, which record high quality audio directly to WAV or MP3 format . In this way, over the years musicians can acquire many hours of field recordings in standard audio formats. Similarly, organisations such as Na Píobairí Uilleann, Comhaltas Ceoltóirí Éireann and the Irish Traditional Music Archive have been acquiring field recordings of traditional music for over sixty years and these organisations now possess many thousands of hours of recordings in a variety of formats and on a variety of different media .

In order for these archives to be usefully searched, they must be annotated with appropriate metadata, such as tune names, time signatures, key signatures and instruments. Additionally for musicological and ethnographic study, archives could be annotated with stylistic metadata. This is the main goal of the MATT2 project and this paper presents a system for automatically annotating field recordings of monophonic traditional dance music. First, using a number of DSP (Digital Signal Processing) algorithms, tunes are transcribed to the ABC music notation language. Once a transcription is made, the system compares it against a corpus of 860 human made tune transcriptions. The ABC language has the advantage of being based on ASCII text and so tunes in ABC can be easily processed and analysed using algorithms for textual information retrieval. Using this approach, a high success rate for both long and short phrases of music is reported. Section 2 of this paper presents a brief discussion of traditional music archiving. Section 3 presents related work on the use of ABC and on MIR (Music Information Retrieval) systems in general. Section 4 describes the algorithms used in implementing MATT2. Section 5 presents experimental results which establish the effectiveness of MATT2 by annotating fifty pieces of input audio recorded in imperfect conditions. Section 6 presents conclusions and future work.

**2. Background**

Current estimates suggest there are at least seven thousand traditional tunes in existence (S Driscoll 2004; Wallis & Wilson 2001). It is proposed that in the past, many isolated rural communities in Ireland developed their own repertoire of tunes and that widespread knowledge of a common repertoire did not occur until the publication of catalogues of traditional tunes such as O’Neill’s The Music of Ireland in 1903 (Krassen 1975; Keegan 1992). In his seminal work, O’Neill collected 1850 tunes played by emigrant Irish musicians in Chicago in the latter part of the nineteenth century.

In 1991, the ABC music notation language was introduced by Chris Walshaw . The format was designed primarily for folk and traditional tunes which can be written on one stave in standard classical notation. ABC files are ASCII text files and so can be edited by any text editor, without the necessity for special software. Each file (known as a *tune book*) can contain multiple tunes. There is an active Internet community engaged in the transcription of tunes in ABC format. For example, the website thesession.org has over seven thousand tunes entered by the traditional music community .

The tune given in Figure 1 is typical of the transcriptions that can be sourced in ABC from publicly available databases. In this transcription the transcriber has included a significant amount of useful metadata with the notation for the tune such as the source of the transcription, the discography and a listing of similar tunes.

X:422  
T:Come West Along the Road  
R:reel  
S:Session  
H:See also #432, in A. This version is also played in A.  
H:1st part similar to "Over the Moor to Peggy", #710  
D:Arcady: Many Happy Returns  
D:Noel Hill & Tony McMahon: \'I gCnoc na Gra\'i  
Z:id:hn-reel-422  
M:C|  
K:G  
d2BG dGBG|~G2Bd efge|d2BG dGBG|1 ABcd edBc:|2 ABcd edBd||  
|:g2bg egdg|(3efg dg edBd|1 g2bg egdB|ABcd edBd:|2 gabg efge|dega bage||

Figure : The tune "Come West Along the Road" in the ABC format [11].

The body encoding supports such features as ornaments, bar divisions, sharps, flats, naturals, repeated sections, key changes, guitar chords, lyrics and variations.

Between 1997 and 2000, a group of musicians under the leadership of Dan Beimborn and John Chambers, undertook a grass roots project to transcribe three of O’Neill’s books to electronic format using the ABC music notation language. As copyright had expired on O’Neill’s original books, they made their work freely available on the Internet .

Many of the tunes from O’Neill’s books are played differently by musicians today, as is normal with a living tradition. Around the same period (the late 1990’s) Henrik Norbeck collected nearly two thousand tunes in ABC format from various sessions and recordings. This collection contains many modern settings of tunes from O’Neill’s books . The MATT2 system uses a corpus of reels in ABC format drawn from Henrik Norbeck’s transcriptions.

**2. Related Work**

In related work, ABC transcriptions have been used to encourage creativity in traditional music sessions and to facilitate e-learning in non traditional settings (Bryan Duggan 2007; Bryan Duggan 2007; Bryan Duggan 2006) by making available thousands of tunes in ABC format on a mobile device. To achieve this, TunePal was developed. TunePal runs on a variety of handheld computing devices running Symbian OS and the Windows Mobile operating system. TunePal supports the retrieval of tunes by text searches and playback of ABC files and is in use by many hundreds of traditional musicians . ABC transcriptions have also been used in two computational creativity applications . To develop MATT1 (Machine Learning Articulation of Traditional Tunes) detailed transcriptions were made of the playing of flute player and maker Eamonn Cotter. These were used to train a machine learning system to *ornament* traditional music in the style of the musician. In a *computational creativity* algorithm is presented that can compose new reels by analysing the structure and n-gram note sequences present in a corpus of reels. A group of domain experts then evaluated the generated tunes for aesthetic value and correctness.

The application proposed in this paper is an MIR system. These systems convert melodies to sequences of symbols that can be matched against query patterns using methods from general string matching. To overcome transcription errors several applications use a pitch contour representation (up, down, or repeat) to give a representation of the direction of the pitch intervals (Ghias et al. 1995; Dannenberg et al. 2004). Many systems also use hummed or sung queries as input (Ghias et al. 1995; N Hu & Dannenberg 2002; Birmingham et al. n.d.). Hu & Dannenberg (N Hu & Dannenberg 2002) compare a number of approaches to MIR and conclude that MIR systems must deal with many difficult problems, including robustness in the face of transcription errors, transposition invariance and tempo variance (*rubatto*). Algorithms must also allow searching for substrings within an overall melody. Widmer *et al*. (Widmer et al. 2005) similarly classify the problem of feature extraction from digital signals for MIR systems as error prone and suggest that systems should be developed that model higher level models of music cognition that human listeners apply when listening to music. They describe *GenreCrawler*, an alternative to DSP approaches that uses web mining as a technique for MIR. MIR in traditional Irish music has an additional difficulty in that traditional musicians rarely play tunes as transcribed in books. In fact it is reported that a good traditional musician will almost never play a tune twice, identically (Larson 2003; Keegan 1992; Hamilton 1990; Tansey 1999; Vallely 2004).

Interestingly, Adams *et al.* (Adams, MA Bartsch & Wakefield 2003) suggest that coarsely quantised melodic representations such as simplified pitch contour representations do not improve retrieval performance for query-by-humming systems. Also we suggest that humming is an inefficient query mechanism for the problem being addressed by this research. A far better approach is to allow queries to be generated directly from the playing of traditional instruments. While there are MIR systems that allow users to search for traditional Irish dance tunes using text based musical queries (Bryan Duggan 2006) and there are MIR systems that allow users to search for melodies using sung or hummed queries (Birmingham et al. n.d.), there are no MIR systems that we are aware of that allow musicians to search for traditional Irish dance tunes using queries played on traditional instruments. Some examples of the above include the website thesession.org which contains an extensive collection of over seven thousand traditional dance tunes in the ABC language. The session.org supports text queries by any of the metadata associated with a tune or melodic queries in the ABC language. Similarly, Melodyhound a publicly accessible MIR system that supports sung queries and contains a large collection of traditional Irish dance tunes does not generate positive results when queries are presented in the form of melodies played on the tin-whistle or wooden flute.

Despite the difficulties outlined, in experiments, MATT2 demonstrates robust transcription accuracy. It also matches audio from substrings in the corpus and so can match any phrase from a tune, not just the incipit. To our knowledge, MATT2 represents the first attempt to adapt MIR to the specific characteristics of traditional Irish dance music and to support queries played on traditional instruments.

MATT2 makes use of the edit distance string matching algorithm. Edit distance also known as *Levenshtein distance* or *evolutionary distance* (Levenshtein 1966; Navarro & Raffinot 2002) is a concept from information retrieval and it describes the number of edits (insertions, deletions and substitutions) that have to be made in order to change one string to another. It is the most common measure to expose the similarity between 2 strings. The edit distance *ed(x, y)* between strings *x=x1 ... xm* and *y=y1 ... yn*, where *x, y* is the minimum cost of a sequence of editing steps required to convert *x* into *y.*  is the alphabet of possible characters and  is the set of all possible sequences of *ch* . Edit distance can be calculated using dynamic programming . Variations on the edit distance algorithm have been applied in domains such as DNA analysis and automated spell checking and are commonly used in MIR systems (N Hu & Dannenberg 2002; Birmingham et al. n.d.; Lemstrom & Perttu 2000).

**3. System Design**

MATT2 works on mono, digital audio files in the WAV format recorded at 44KHz. A high level diagram of the sub systems that make up MATT2 are presented in . The subsystems present in MATT2 will now be described.

**3.1. Onset Detection**

The audio file to be annotated is first segmented into candidate note onsets and offsets using an onset detection function adapted from Gainza . The onset detection function ODCF is based on time domain FIR comb filters. ODCF discovers harmonic characteristics of the input signal and is therefore tolerant to energy changes in an input signal not caused by note onsets and is also better at detecting onsets in *legato* playing typical of windblown traditional instruments such as the flute and the tin whistle.

The input signal is first segmented into overlapping frames of 2048 samples (approximately 46 milliseconds). Each frame overlaps with the previous frame by 75%. Each frame is then passed through a bank of twelve FIR comb filters. A FIR comb filter works by summing the input signal with a delayed version of the same input signal. The delay of the filter is calculated as being the length in time of a single period of a waveform at the frequency.



Figure : High level diagram of the MATT2 tune annotation system

This has the effect of amplifying the frequency or a harmonic thereof in the input signal that matches the frequency being filtered. Thus, the energy of the input signal is doubled only if the peaks of the signal coincide with the peaks of the FIR comb filter. This will only occur for a given delay and its integer multiples . Twelve filters *E(m, D)* with different delays are used corresponding to the twelve semitones in the key of D3 as per (1).

|  |  |
| --- | --- |
|  | (1) (M. Gainza 2006) |

For each frame of audio examined, the outputs of the audio passed through each of the twelve filters are calculated. A value for the ODF *dE(m)* is then calculated as being the sum of the difference between the outputs of each of the twelve filters in successive frames squared, as described in (2).

|  |  |
| --- | --- |
|  | (2) (M. Gainza 2006) |

In the case where the pitch of the input signal changes from one note to another, this will result in a peak in the ODF graph. Using statistical techniques, a threshold is calculated for each 25ms of audio as being the overall average ODF plus the standard deviation of the ODF in that frame . Peaks above the threshold are recognised by the system as candidate note onsets. A *peak* in the ODF is defined as a value preceded by four ascending values and followed by four descending values (though MATT2 supports a configurable value for this). Onsets and offsets are considered by the system to be concurrent as the wooden flute is typically played legato and so a candidate note is considered to be a segment bounded by two adjacent onsets.

Figure 3 shows the signal for the first bar of the tune “The Boyne Hunt” with the detected candidate note onsets marked. The second plot in this figure shows the ODF for the signal, with the dynamic threshold and the candidate onsets marked. In this plot, it is significant that the first note contains a dynamic energy change approximately half way through the note which the ODF has correctly ignored.

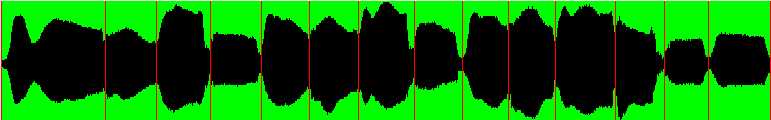
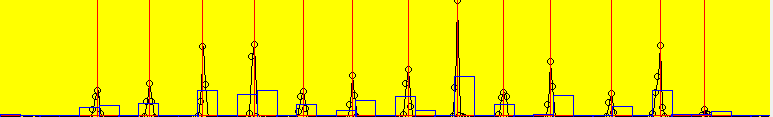
  


Figure : Signal and ODF plots of the first bar of the tune "The Boyne Hunt"

**3.2. Pitch detection**

To establish the perceived pitch of each note, the *fundamental frequency (F0)* of the note must be derived. The pitch detector subsystem first calculates the highest, nearest power of two *nP(m)* for the length in samples *dS(m*) of each segment *m* of audio bordered by a candidate onsetas given in (3).

|  |  |
| --- | --- |
|  | (3) |

It then performs a short term FFT (Fast Fourier Transform) on the segment. To determine the perceived pitch the system first calculates peaks in the FFT. In this case, a peak is a value bordered by two ascending/descending values. The algorithm then calculates the interval between the first and second peak. This approach works well for the harmonics of the wooden flute.

**5.1. Breath detection**

The flute is a windblown instrument and hence a musician must periodically take breaths as a piece of music is being played . MATT2 incorporates a breath detector sub system to transcribe a breath in the signal. The breath detector first calculates the average absolute amplitude of the entire signal *x(m)*. It then calculates the average absolute amplitude for each candidate note frame *x(n)*.

A breath is marked if either the perceived pitch detected by the pitch detector is less than 100Hz or the average amplitude of a candidate note *n* is less than a 10% threshold *t* of the average amplitude over the entire signal *m* asin (4). Again, this threshold is configurable. Breaths detected before the transcription of the first pitched note are ignored by the system.

|  |  |
| --- | --- |
|  | (4) |

**3.3. Pitch spelling**

A wooden flute used to play traditional music has a range of two octaves, though this can be extended by cross fingering techniques . To tag each candidate note with a pitch spelling *pS(n)*, each calculated note frequency is compared with the frequencies of the notes in the key of D3 Major and D4 Major *k1... k16* the two octaves playable on a wooden flute. The nearest match for the frequency *f(n)* is the assigned the pitch spelling *pS(n)* (5).

|  |  |
| --- | --- |
|  | (5) |

**3.4. Note quantisation**

At this stage, the system has a set of candidate notes, with corresponding durations and perceived pitches with spellings. To quantise note durations, the quantisation subsystem generates a *fuzzy histogram* of the intervals between consecutive onsets. As the test matching corpus contains only reels (tunes in 4/4 time) the histogram bin with the highest count is considered to be the length of a quaver note. The fuzzy histogram counts notes within 30% +/- of the the bin width. This 30% is the *fuzz* referred to in . The algorithm also updates the bin width each time a candidate is added to the bin, so that the bin widths contain the cumulative average lengths of notes counted.

Using the histogram, the duration calculator concludes the length of a quaver *qL* as being the bin with the highest value. Pseudocode for this algorithm is given in .

Figure 5 shows the duration histogram for a twenty eight second phrase of music from the tune “The Hunters Purse”. Notes whose durations are close to zero have their durations merged with subsequent notes. This has the effect of eliminating consecutive onsets (false positives caused by noisy onsets) and also eliminating ornamentation notes such as those found in *rolls*, *cuts* *taps* and *crans* typical of traditional Irish music .

MATT2 then calculates the maximum bin value in a second fuzzy histogram of the new note durations after consecutive onset and ornamentation elimination. The system uses this value to be the new length of a quaver *qL’*. The duration calculator then evaluates the nearest multiple *qQ* of the quaver length *qL* for each candidate note *n* as per (6).

Figure : *Fuzzy histogram* of candidate note lengths from a 28 second phrase from the tune "The Hunters Purse"

|  |  |
| --- | --- |
|  | (6) |

In this way notes are quantised as being quavers, crochets, dotted crochets and minims and ornamentation notes are eliminated from the transcription.

*foreach* (note *in* transcribed\_notes)

*begin*

found 🡨 false

*foreach*(bin *in* histogram)

*begin*

bin\_start 🡨 bin.width - fuzz

bin\_end 🡨 bin.width + fuzz

*if* (note.duration >= bin\_start *and* note.duration <= bin\_end)

*begin*

found 🡨 true

bin.count ++

bin.width 🡨 (bin.width +  
 note.duration) / 2

*break*

*end*

*if not* found

*begin*

newNote.count 🡨 1

newNote.width 🡨 note.duration

histogram.add(newNote)

*end*

*end*

quaver\_length 🡨 max(histogram)

Figure : Pseudocode for the fuzzy histogram quaver length calculator

For many of the test recordings used to evaluate MATT2 recorded in imperfect conditions, this approach results in remarkably few transcription errors.

**3.5. ABC Normalisation**

Before edit distance matching against the corpus is carried out, both the transcribed string and strings from the corpus are normalised. This step is necessary as the ABC format supports features such as repeated sections, which need to be expanded so that they can be correctly matched against transcribed phrases. Normalisation of musical strings has the added advantage of minimising the effect of transcription errors on the calculation of the edit distance. Normalisation involves four stages.

Firstly, all whitespace, ornamentation markers and text comments are removed. When ornamentation markers (*~{}*) are removed from ABC transcriptions, this has the effect of quantising the duration of the majority of notes in corpus strings to multiples of the duration of a quaver.

Original:

**d2BG dGBG|~G2Bd efge|d2BG dGBG|1 ABcd edBc:|2 ABcd edBd||**

After Ornamentation removal:

**d2BGdGBG|G2Bdefge|d2BGdGBG|1ABcd edBc:|2ABcdedBd||**

After note expansion:

**ddBGdGBG|GGBdefge|ddBGdGBG|1ABcd edBc:|2ABcdedBd||**

After section expansion:

**ddBGdGBGGGBdefgeddBGdGBGABcdedBc**

**ddBGdGBGGGBdefgeddBGdGBGABcdedBd**

After register normalisation:

**ddBGdGBGGGBdefgeddBGdGBGABcdedBc**

**ddBGdGBGGGBdefgeddBGdGBGABcdedBd**

Figure : Normalisation stages for the A part of the tune “Come West Along the Road”

Secondly, all notes of duration greater than that of a quaver are expanded to be multiple instances of a quaver. This minimises the effect of false negatives in the ODF as *long* notes (false negatives in the ODF) become multiple short notes in both the transcribed phrase and in tunes from the corpus . This also introduces a certain amount of tolerance in the matching subsystem to stylistic variations in the playing of tunes.

Thirdly repeated sections are expanded and bar divisions are removed. ABC supports several notations for different types of repeated phrases . This means for example, that if the transcribed tune was the A part of a tune played twice, this would be correctly matched against the expanded A part of a tune from the corpus.

Finally all notes are transformed to be in the same register. This is achieved by transforming lower case characters in the ABC of tunes to upper case. shows examples of each stage in the ABC normalisation process.

**3.6. Edit distance matching**

One final transformation is carried out on strings from the corpus before they are compared with transcribed strings. Occasionally, strings from the corpus are shorter than transcribed strings. For example, the transcribed string might be from a double reel, while the string from the corpus could be from a single reel (a tune half the length). In order to gain the maximal impact from the transcription, corpus strings shorter then transcribed strings are duplicated until their length is greater than the length of the transcribed string. This approximates what a real musician would do in order to extend the duration of a tune .

The minimum edit distance *eF(c)* for each string *c* from the corpus *Z* then calculated using a cost of one for insertions, deletions and substitutions, for each pair consisting of the transcribed string *s* in substrings of *c*. A variation of the classic edit distance algorithm described in is used to search for the minimum edit distance for a search string in substrings of a target string. This way any phrase from a tune can be matched not just complete tunes and not just incipits. Edit distances are normalised by dividing by the length of the transcribed string to produce *eF(c)*. Two methods for establishing the lowest edit distance were implemented. Firstly, MATT2 returns tunes whose edit distance is less than a configurable threshold. The system also returns the top ten matching tunes in order of lowest edit distance.

**4. Interface**

MATT2 was developed in Java. A screenshot of the system is presented in .

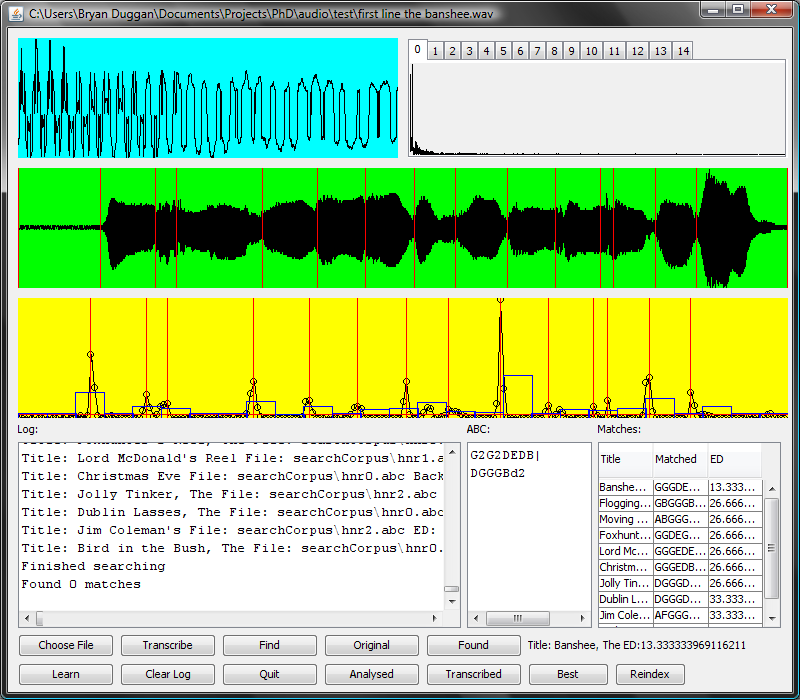


Figure : Screenshot of MATT2

The interface to MATT2 displays several useful plots of the outputs of each stage in the transcription and the matching algorithm such as the current frame being analysed, the onset detection function and the FFT of each detected note. Additionally, the interface displays the transcription in ABC format and the title of the current closest tune match. MATT2 can also play the original WAV file being analysed, the transcribed pitches and durations, the quantised transcription in ABC, the closest match and any of the matched tunes. When the matching algorithm terminates, MATT2 displays and can play any of the top ten closest matching tunes, with their corresponding edit distances. It can also operate in batch mode where it will attempt to annotate all the WAV files in a folder.

**4. Experiment and results**

To evaluate MATT2, nine subjects recorded the A and B parts to a number of double and single reels. These recordings were made in imperfect conditions (a kitchen in a house, a school room and a pub) and contain ambient noise such as chairs moving, doors opening and foot taps. The recordings were edited so that the audio being tested contained a mix of complete tunes and segments of tunes taken from the start, middle and end of tunes. Some deliberately challenging audio such as archive recordings, flute duets**,** flute and fiddle duets, and fiddle solos was also included.

For forty three of input audio files, the algorithm correctly identified the tune. A further five tunes were correctly annotated within the top ten closest matches. Just two of the test audio files were incorrectly annotated. Experiments have shown that MATT2 demonstrates a remarkable robustness to tempo variations, musical style, ornamentation, musician variations, instruments and recording environments. Amazingly, one of the tunes tested was recorded over thirty years ago by the flute player Packie Duignan on an analogue tape recorder and the audio had badly degraded. We did not attempt to pre-process this audio before annotating it. For this test, the algorithm did not correctly identify the tune; however the correct tune was the second closest match. The correct tune could have been returned with the addition of a simple heuristic. Further fiddle solos, flute duets and flute and fiddle duet was also included in test audio. In all cases the algorithm correctly identified the tune, despite the fact that the onset detection function used is reported to be inaccurate at detecting onsets in fiddle recordings .

The average normalised edit distance was 39% for the closest matches, while the average of the nearest second match was 53%. The difference between these (14%) can be considered as a confidence level. As an example of the results the system generates, shows the edit distances returned by MATT2 for the tune, “The Golden Keyboard”, played by flute player Eamonn Cotter. This recording was made on a portable MP3 recorder and subsequently transcoded to WAV format.

Figure : Top ten edit distances for a recording of the tune "The Golden Keyboard"

The edit distances shown in the graph are normalised and there is a 14% difference between the correct tune and the nearest closest match. The second and subsequent closest matches have only small variances. These results are summarised in .

|  |  |
| --- | --- |
| Average ED (closest): | 39% |
| Average ED (next): | 53% |
| Difference: | 14% |
| Average ED (closest, correct): | 36% |
| Average ED (next, correct) | 52% |
| Difference: | 16% |
| Average ED (closest, incorrect): | 56% |
| Average ED (next, incorrect): | 57% |
| Difference: | 1% |

Table : Average edit distances for the closest match and the next closest match for tunes correctly and incorrectly annotated

When incorrect matches were considered, it was discovered that in all cases incorrect matches were as a result of transcription errors. Further, the transcription errors were caused by either unusually prominent foot taps or the musician started slow and then speeded up as the tune was played, which affected note duration quantisation.

**6. CONCLusions**

In this paper MATT2 a system for annotating field recordings of traditional Irish music with metadata was described. MATT2 combines a novel transcription system that makes use of ODCF to detect onsets and fuzzy histograms to quantise note durations. MATT2 also makes use of publicly available transcriptions in ABC made by the traditional music community to match transcriptions against. A novel string normalisation technique that takes advantage of the ABC language to eliminate the effect of transcription errors and stylistic variations in input audio was also presented. Further MATT2 improves on pitch contour representations of music strings in MIR systems by using accurate pitch and duration information from input audio in the melody matching subsystem. To our knowledge, MATT2 represents the first attempt to adapt MIR to the specific characteristics of traditional Irish dance music.

Experiments demonstrate that the approach outlined in this paper is robust to variations in musician, style and instrument. In successfully testing MATT2 on badly degraded archive audio, we conclude that the approach presented can be further developed for use on the many thousands of hours of archived recordings of traditional music that currently exist and that are being collected. Our system currently works on segments from tunes or complete tunes. One interesting complication we have not addressed is that tunes are usually played in sets of two or more tunes and each tune is usually played twice. We therefore have the problem that a single audio input file may contain several tunes.

In the cases where the algorithm did not correctly identify the correct tune, we conclude that the transcription subsystem was not able to accurately transcribe the tune. We therefore feel that improving transcription accuracy will lead to more accurate matching and will focus on this problem in future work. Widmer *et al*. (Widmer et al. 2005) state that transcription algorithms need the kind of higher level musical knowledge that humans poses and we hope to try and develop this approach, possibly using a *preference rule* style approach similar to the work of Temperly . Support for the transcription of tunes in different time signatures is planned as is experimentation with corpora of tunes in other popular time signatures such as jigs and hornpipes.

1. Machine Annotation of Traditional Sets (MATS)

Several papers address the necessity of developing MIR (Music Information Retrieval) systems that are adapted to the specific requirements of ethnic music and also to the needs of musicologists studying ethnic music . While there are MIR systems that allow users to search for traditional Irish dance tunes using text based musical queries and there are MIR systems that allow users to search for melodies using sung queries , there are no MIR systems that we are aware of that allow musicians to search for traditional Irish dance tunes using queries played on traditional instruments. Some examples of the above include the website thesession.org which contains an extensive collection of over seven thousand traditional dance tunes in the ABC language; the system supports text queries by any of the metadata associated with a tune or melodic queries in the ABC language. Similarly, Melodyhound a publicly accessible MIR system that supports sung queries and contains a large collection of traditional Irish dance tunes does not generate positive results when queries are presented in the form of melodies played on the tin-whistle or wooden flute.

Such a system would have many applications in the field of music archiving and retrieval, particularly given the many thousands of hours of archive music collected by organisations involved in the cataloguing of traditional music such as Na Píobairí Uilleann, Comhaltas Ceoltóirí Éireann and the Irish Traditional Music Archive. Similarly it is common at traditional music sessions, recitals and even on commercial recordings for tunes to be named *gan ainm* (without name) when the tune in question does in fact have a name, composer and history. For a typical example see the CD recording .

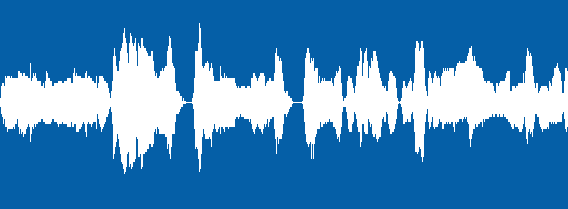
Previous work proposes MATT2 (Machine Annotation of Traditional Tunes) as a system that can identify tunes played on either the flute or the tin whistle . MATT2 takes advantage of a number of novel subsystems that significantly increase matching accuracy for traditional tunes played in a variety of regional styles by different musicians. These include an onset detection function developed for windblown instruments, an ornamentation compensation algorithm based on fuzzy histograms, a two thousand tune corpus of tunes in the ABC language (a natural fit for traditional music) and a melody normalisation algorithm that adapts tunes in the corpus to the way they might be played by a human musician. MATT2 is described in detail in and we present an overview in section 3. The main purpose of this paper is to present our enhancements to the MATT2 system and specifically to present a new algorithm for annotating sets of traditional Irish dance tunes. Previous versions of MATT2 could only annotate single tunes, however in traditional music tunes are rarely played singly. More commonly tunes are played in groups of at least two tunes known as a *set* of tunes. A set typically consists of two three or four tunes played in succession without an interval (Vallely 1999; Bryan Duggan, O'Shea & Padraic Cunningham 2008). Typically each tune in the set is played twice or three times before musicians advance to the subsequent tune in the set. A repetition or a change from one tune to the next in a set is known as a *turn*. As tunes in sets are always in the same time signature and often in the same key, the challenge therefore is in segmenting sets into tunes and repetitions. The approach presented in this paper tackles this problem by making use of melodic similarity calculated using a variant of the *edit distance* string matching algorithm described in section 3. The MATS algorithm described in this paper can identify the start and end of each repetition of a tune, can count the repetitions and can identify the title and associated metadata associated with each tune in a set.

Section 2 of this paper briefly explains the domain of traditional Irish dance music. In Section 3 existing work on the MATT2 system is presented. Section 4 presents MATS (Machine Annotation of Traditional Sets), a novel annotation algorithm which annotates sets of traditional tunes. Section 5 presents experimental results which establish the effectiveness of this new algorithm and section 6 presents conclusions and future work.

**2. traditional irish dance music**

The most common forms of dance music are *reels, double jigs* and *hornpipes*. Other tune types include *marches*, *set dances*, *polkas, mazurkas, slip jigs, single jigs and reels, flings, highlands, scottisches, barn dances, strathspeys* and *waltzes* . These forms differ in time signature, tempo and structure. For example a reel is generally played at a lively tempo and is in 4/4 time (four crochets in a bar, though usually transcribed as eight quavers in a bar), while a waltz is generally played at slower pace and is in 3/4 time. Most tunes consist of a common structure of two parts called the *A* part and *B* part. Tunes are typically played as *sets*. Certain common sets were originally put together to accompany set dances (Vallely 1999), while other sets have become popular as a result of being recorded by emigrant Irish musicians in America in the early part of the twentieth century.

turn



Time

Figure : Waveform of the last phrase from the tune "Jim Coleman’s" and the first phrase from the tune "George Whites Favourite" played in a set

The origin of many sets of tunes is unknown and musicians often compile new sets “on the fly” in traditional music sessions. Figure 15 shows a waveform plot from two tunes played in a set. The tunes were played on a wooden flute and as can be seen in the plot, there is no interval between the end of the first tune and the start of the second tune. Maddage *et al.* and other segmentation approaches generally look for repetitive patters in a music recording . This is not the case in our approach, where each tune in the set can be played once or many times.

When a traditional musician plays a tune, it is rarely played exactly as transcribed. In fact an experienced musician never plays the same tune twice identically, employing the subtleties of *ornamentation* and *variation* to interpret the tune . For a discussion on the use of ornamentation in traditional music we refer to .

Ornamentation plays a key role in the individual interpretation of traditional Irish music . The usage of ornamentation is highly personal and large variations exist in the employment of ornamentation from region to region, instrument to instrument and from musician to musician. Tansey colourfully describes ornamentation in the following way:

“*I put it to you therefore that it had to come from the throats of birds, the wild animals, the ancient chants of our forefathers, the hum of the bees and the mighty rhythms of the galloping hooves of wild horses all moulded together…*”

Ornamentation is difficult to detect correctly and state of the art ornamentation detection algorithms report a success rate of just 40% for multi-note ornaments (M. Gainza 2006; M Gainza & Coyle 2007). Similarly, related work in classical music suggests that the playing of ornamentation (grace notes) requires adaptation of melodic similarity measures .

It is clear from this brief introduction that an MIR system for traditional dance music must therefore deal with many special problems, such as stylistic variation even within the same instance of a tune, the use of ornamentation which can skew melodic similarity measures and the collection of tunes into sets creating segmentation problems. Transposition invariance is not a requirement for MIR in traditional music as it is uncommon for tunes to be transposed into different keys .

**3. machine annotation of traditional tunes (MATT2)**

MATT2 works on mono, digital audio files in the WAV format recorded at 44KHz. A high level diagram of the subsystems that make up MATT2 are presented in Figure 16. MATT2 is described in detail in and so a brief description is presented here.

 Figure : High level diagram of the MATT2 tune annotator

The audio file to be annotated is first segmented into candidate note onsets using an onset detection function adapted from Gainza (M. Gainza 2006; M Gainza & Coyle 2007). The onset detection function ODCF is based on time domain FIR (Finite Impulse Response) comb filters. ODCF discovers harmonic characteristics of the input signal and is therefore useful for detecting onsets in *legato* playing typical of windblown traditional instruments such as the flute and the tin whistle.

In order to detect the perceived pitch of a frame, the pitch detection sub-system performs a STFT (Short Term Fast Fourier Transform) on segments bounded by onsets detected by the onset detection system. The algorithm then calculates the pitch as being the interval between the two most prominent peaks in the FFT graph. This simple approach works well for the harmonics of the wooden flute and the tin whistle.

MATT2 incorporates a breath detector subsystem to transcribe a breath in the signal. A breath is marked if either the pitch detected by the pitch detector is less than 100Hz or the average amplitude of a candidate note c*n* is less than a 10% threshold *th* of the average amplitude of the entire signal *s*. Breaths detected before the transcription of the first pitched note and at the end of the transcription are ignored by the system.

MATT2 uses a heuristic to determine if the input signal was generated by a tin whistle or a wooden flute. A tin whistle in the key of D is pitched exactly one octave above a flute in the key of D, so if the algorithm counts more notes with a pitch above G5 (783.99hz) than below G5, then the algorithm concludes that the input signal contains a tin whistle and the pitches in the pitch spelling algorithm are shifted up accordingly.

Both the wooden flute and the tin whistle have a range of two octaves, though this can be extended by cross fingering techniques (Larson 2003; Hamilton 1990; Vallely n.d.). To tag each candidate note *cn* with a pitch spelling *pS(cn)*, each calculated note frequency is compared with the frequencies of the notes in the key of D4 Major and D5 Major *k1... k16* the two octaves playable on a wooden flute.

The systemeliminates notes whose durations are close to zero by merging their durations with subsequent notes. This has the effect of eliminating consecutive onsets (false positives in the ODF caused by noisy onsets) and also eliminating ornamentation notes such as those found in *rolls*, *cuts* *taps* and *crans* typical of traditional Irish music . To achieve this, the quantisation subsystem first generates a histogram of all the note durations. The histogram bin with the highest value is considered to be the length of a quaver note. The histogram counts notes within +/-30% of the bin width. The algorithm also updates the bin width each time a candidate is counted, so that the bin widths represent the cumulative average lengths of notes counted. A transcription *t* is then generated in the ABC language of the input signal from the features extracted by the subsystems in MATT2.

MATT2 has a corpus *Z* of two thousand known tunes (and variations) in the ABC language drawn from the collections of Norbeck . To identify a tune, MATT2 firstly normalises both the transcription *t* and each string *c* *Z*. This process is described in detail in . Normalisation minimises the effect of transcription errors and stylistic variation on the calculation of melodic similarity. The *edit distance* is then calculated for *t* in every *c* *Z* and the tune with the lowest edit distance is returned as a match.

Edit distance, also known as *Levenshtein distance* or *evolutionary distance* (Levenshtein 1966; Navarro & Raffinot 2002), is a concept from information retrieval and it describes the number of edits (insertions, deletions and substitutions) that have to be made in order to change one string to another. It is the most common measure to expose the similarity between 2 strings.

The edit distance *ed(x, y)* between strings *x=x1 ... xm* and *y=y1 ... yn*, where *x, y* is the minimum cost of a sequence of editing steps required to convert *x* into *y.*  is the alphabet of possible characters and  is the set of all possible sequences of *ch* . Edit distance can be calculated using dynamic programming . To compute the edit distance *ed(x,y)* a matrix *M1...m+1,1...n+1* is constructed where *Mi,j* is the minimum number of edit operations needed to match *x1...i* to *y1...j*. Each matrix element *Mi,j*  is calculated as per (1). The minimum edit distance between *x* and *y* is given by the matrix entry at position *Mm+1,n+1*.

|  |  |  |
| --- | --- | --- |
|  | if *xi=yi*  else | (1) |

The algorithm can be adapted to find the lowest edit distances for *x* in substrings of *y*. This is achieved by setting *M1,j* = 0 for all *j* .*n+1.* In contrast to the edit distance algorithm described above, the last row *Mm+1,j* is then used to give a *sliding window* edit distance for *x* in substrings of *y* (Navarro & Raffinot 2002)*.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | D | G | G | G | D | G | B | D | E | F | G | A | B |
|  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| D | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 2 | 2 | 2 | 1 |
| E | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 1 | 0 | 1 | 2 | 3 | 2 |
| E | 4 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 2 | 1 | 1 | 2 | 3 | 3 |

Table : Edit distance for the string BDEE in DGGGDGBDEFGAB. This string represents the first 13 notes from the tune "Jim Coleman's" in normalised ABC format

An example of this variation on the edit distance applied to search for the pattern “BDEE” in “DGGGDGBDEFGAB” is given in Table 2. The minimum edit distance positions are highlighted. Variations on the edit distance algorithm have been applied in domains such as DNA analysis and automated spell checking and are commonly used in MIR systems (Birmingham et al. n.d.; Lemstrom & Perttu 2000).

With test input drawn from the playing of ten different musicians playing flute, whistle and fiddle, the system was able to correctly identify the tune in 86% of cases. In 96% of cases, the correct tune was identified within the top five closest matches .

**4. MACHINE annotation of TRADITIONAL sets algorithm (MATS)**

In this section MATS is described. MATS is an enhancement to MATT2 described in the previous section. The purpose of MATS is to annotate tunes played in sets. The shortest tune in the corpus *Z* used by MATT2 is a single jig. A single jig *sj* is a tune in 6/8 time with an A and B part played singly (48 quaver notes in duration). The shortest possible set therefore would contain two single jigs (96 notes) played with no repetitions. To annotate a set of tunes, MATS first uses a heuristic to determine if the string of transcribed notes *t* is longer than the length of the shortest set *length(sj)×2*.

When this is the case, the MATS algorithm is used instead of the minimum edit distance algorithm described in section 3. Pseudocode for the MATS algorithm is presented in Figure 17.

MATS first extracts a substring *ss* from *t* the transcription such that *length(ss) = length(sj)* at position *p=1* in *t*. MATS then searches the corpus *Z* using the edit distance algorithm described in section 3 to find a the closest match for *ss*. When a match is found MATS knows the name of the first tune and has *c'*, a transcription of the tune played with no repetitions from the corpus Z. MATS then generates an edit distance profile *edp* for *c',* the matching tune, in *t* the transcription*.* *edp* is given as the last row of the edit distance matrix and can be understood as the positions where substrings in *t* match *c'* withthe minimum edit distance *.*

Figure 18shows the edit distance profiles for the set of tunes “Jim Coleman’s”, “George Whites Favourite” and “the Virginia” played in a set. The algorithm has identified the first tune as “Jim Coleman’s” and has subsequently generated an edit distance profile (the top plot in Figure 18) for the first tune in the transcription. The two troughs in this graph indicate the end of the two repetitions of the tune in the transcription. These can be considered as turns in the set.

The MATS algorithm then normalises the edit distance profile *edp* and passes the graph through a low pass filter that filters frequencies less than 10Hz. This has the effect of smoothing the graph. An example of a smoothed edit distance profile is given in Figure 19. This graph illustrates the top graph in Figure 18 after filtering has been applied.

The algorithm then detects troughs in the graph less than a threshold initially set to *t=0.3*. The algorithm varies this threshold dynamically by trying different values until the number of troughs in the graph is between one and five. It is rare in traditional music for a tune to be repeated more than five times in a set.

p 🡨 0

rem 🡨 length(t) - p

*while* (rem >= sj)

*begin*

ss 🡨 substring(t, p, p + sj)

*foreach* (c *in* Z)

*begin*

ed\_c 🡨 min(ed(ss, c))

*if* (ed\_c < min\_ed)

*begin*

min\_ed 🡨 ed\_c

c' 🡨 c

*end*

*end*

edp 🡨 ed(c', t)

edp 🡨 normalise(edp)

edp 🡨 filter(edp, 10)

th 🡨 0.3

v 🡨 troughs(edp, th)

*foreach* (tr *in* v)

*begin*

convertToTime(tr)

*end*

r 🡨 length(v)

p 🡨 v[r]

print c’, r

rem 🡨 length(t) - p

*end*

Figure : Pseudocode for the MATS set annotation algorithm

The trough detection algorithm in MATS returns a vector of troughs , such that *length()* is the number of troughs and the elements in are the positions of the bottom of the troughs. A trough in MATS need only have a descending wall as a trough can occur at the end of a tune and hence may not contain an ascending wall. An example of this is the third plot in Figure 18.

The algorithm repeats this process with a new *p* given as the last entry in the troughs vector to extract the second and subsequent tunes in the set until it is no longer possible to extract a substring *ss* of length *length(sj)* starting at *p* because we have reached the end of *t*. The second tune in the set, “George Whites Favourite” was played once and there is a corresponding single trough in the graph of the edit distance function (the middle plot in Figure 18) for the tune from the corpus *c'* in the transcription *t*. The third tune “the Virginia” was repeated twice and so there are two troughs in the bottom plot in Figure 18.

**5. RESULTS**

In order to test the robustness of MATS we had a traditional musician record ten audio files of flute tunes played in sets. The recorded files are available at http://www.comp.dit.ie/bduggan/mats. The sets played in the input audio were taken from the Foinn Seisiún series of books published by Comhaltas Ceoltóirí Éireann (Brian 2004).



Figure : Edit distance profiles for three tunes played in a set

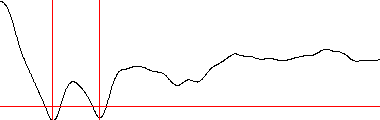




Figure : Filtered version of first graph in Figure 4. The dynamic threshold and detected troughs are marked

The sets consisted of single and double jigs and reels played multiple times in sets. In total, the sets contained 23 separate tunes with 48 turns we were interested in annotating. In carrying out this experiment, we were interested in establishing if MATT2 could correctly figure out the timings of turns and could identify the names of the tunes.

|  |  |
| --- | --- |
| **Correctly identified** | 96% |
| **Incorrectly identified** | 4% |

Table : Correctly and incorrectly identified tunes

MATT2 successfully identified 22 out of the 23 tunes, and recognised each input audio file as a set and so used the MATS set annotation algorithm (Table 5).

Table 6 shows a sample of the data collected in this experiment for the audio file used to generate Figure 18 and Figure 19. To establish a ground truth for the experiment, a human domain expert manually annotated the turns in the sets of tunes. In the human and machine columns are listed the onset time for turns in the set. Onset times for changes from one tune to the next are highlighted. From this table it can be seen that on average MATS was within .85 seconds of the human annotations.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tune** | **Human** | **Machine** | **Difference** |
| **1** | 20.68 | 21.10 | 0.43 |
| **1** | 41.42 | 41.9 | 0.48 |
| **2** | 82.72 | 83.15 | 0.43 |
| **3** | 123.88 | 124.44 | 0.56 |
| **3** | 164.49 | 166.85 | 2.36 |
| **Average** |  |  | 0.85 |

Table : Human & machine annotated turns

The overall annotation accuracy is obtained by calculating two different measures *precision* and *recall.* The value of *precision* is calculated as per (2) where *TP* and *FP* are the true positives (correctly identified turns) and false positives (incorrectly identified turns). *recall* is calculated as per equation (3) where *FN* is the number of false negatives (turns in the input signal not detected by the algorithm).

|  |  |
| --- | --- |
|  | (2) |
|  | (3) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TP** | **FN** | **FP** | **precision(%)** | **recall(%)** |
| 39 | 9 | 6 | 87% | 81% |

Table : Annotation accuracy

Table 7 shows the annotation accuracy. In can be seen from *precision* and *recall* that the algorithm provides a high degree of accuracy at detecting turns. Because the algorithm can successfully identify turns, it can also correctly extract a suitable prefix from the subsequent tune in the set and so can identify the tune. *FN*’s were caused by the algorithm failing to correctly identify the transitions between tunes in a set. When this happens the algorithm cannot extract a representative prefix from the next tune and so all subsequent turns are usually misidentified. In some cases, *FP*’s were within a few seconds of the two second threshold we had set.

**6. conclusions & future work**

This paper presented a novel algorithm that addresses a problem in the domain of Irish traditional dance music, that of annotating sets of tunes. As a set can contain an arbitrary number of tunes played segue without an interval and as tunes in sets are repeated an arbitrary number of times, are always in the same time signature and often in the same key, the significant challenge in this problem is in recognising where one tune ends and the next tune starts. The results presented prove that MATS is effective at segmenting sets, counting repetitions and at annotating individual tunes played in a set. To our knowledge this is the first time this problem has been addressed in an MIR system and we suggest that the proposed approach can be adapted to segmenting repeated tunes from other genres played in a segue.

The corpus used currently contains reels and jigs and in future work it will be augmented with the full complement of traditional tunes in different time signatures. One interesting feature not yet exploited is the metadata typically present in an ABC transcription. Effectively the time and key signature of an input audio file can be determined by *melodic similarity* with a known tune. This can be exploited in several interesting ways. Firstly, if the first tune in a set were to be identified as a reel, the search for subsequent tunes can be limited to reels, thus speeding up annotation. Conversely, if a number of reels were to be identified in a set and a single tune in a different time signature was to be identified this could be recognised as a potential error.

1. Conclusions

The following are the list of publications from research carried out in the pursuit of this PhD.

Duggan, B: TunePal: A Portable Tune Teaching Tool for Traditional Musicians, DIT Annual Showcase of Learning & Teaching Activities, January, 2007

Duggan, B., Zheng, C., Cunningham, P.: MATT - A System for Modelling Creativity in Traditional Irish Flute Playing, Third Joint Workshop on Computational Creativity, ECAI'06, Italy, August 2006

Duggan, B.: Learning Traditional Irish Music using a PDA, IADIS Mobile Learning Conference, Trinity College, Dublin, Ireland, July 2006

* 1. Conclusions

Bibliography

Baroni, M. (2006). Hearing Musical Style: Cognative and Creative Problems. Musical Creativity Multidisciplinary Research in Theory and Practice. I. Deliége and G. A. Wiggins, Psychology Press.

Boden, M. A. (1996). Dimensions of creativity. Cambridge, Massachusetts, MIT Press.

Breathnach, B. (1977). Folk Music and Dances of Ireland.

Brown, J. C. (1992). "Musical fundamental frequency tracking using a pattern recognition method." JASA 1992 **3**: 1394-1402.

Canainn, T. O. (1978). Traditional Music in Ireland. London, Routledge and Keegan-Paul Ltd.

Cotter, E. (2006). Interview notes.

Coyle, L., D. Doyle, et al. (2004). "Representing Similarity for CBR in XML, Department of Computer Science, Trinity College Dublin."

Coyle, L., C. Hayes, et al. (2002). "Representing Cases for CBR in XML."

Cristianini, N. and J. Shawe-Taylor (2000). An Introduction to Support Vector Machines and other Kernel-based Learning Methods, Cambridge University Press.

Csikszentmihalyi, M. (1999). Implications of a

Systems Perspective for the Study of Creativity. Handbook of creativity, Cambridge University

Press**:** 313–335.

Dannenberg, R. B., B. Thom, et al. (1997). A Machine Learning Approach to Musical Style Recognition. 1997 International Computer Music Conference, International Computer Music Association.

DeManteras, R. L. and J. L. Arcos (2002). "AI and Music: From Composition to Expressive Performance." Ai Magazine.

Dixon, S. (2004). "On the analysis of musical expression in audio signals."

Dixon, S., W. Goebl, et al. (2002). Real time tracking and visualisation of musical expression. Music and Artificial Intelligence: Second International Conference, ICMAI2002, Edinburgh, Scotland.

Duggan, B., Z. Cui, et al. (2006). MATT - A System for Modelling Creativity in Traditional Irish Flute Playing. Third ECAI Workshop on Computational Creativity, Riva Del Garda, Italy.

Eck, D. and J. Lapamle (2006). Learning Musical Structure Directly from Sequences of Music, University of Montreal

Department of Computer Science

CP 6128, Succ. Centre-Ville

Montreal, Quebec H3C 3J7 Canada.

Furnkranz, J. (1999). "Separate-and-conquer rule learning." Articial Intelligence Review **13**(1): 3-54.

Gainza, M. (2006). "Automating Ornamentation Transcription."

Gainza, M. (2006). Music Transcription within Irish Traditional Music. Faculty of Engineering. Dublin, Dublin Institute of Technology. **PhD**.

Gainza, M., E. Coyle, et al. (2005). Onset Detection Using Comb Filters. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY.

Götz, I. (1981). "On Defining Creativity." Journal of Aesthetics and Art Critism(39): 297-301.

Hamilton, C. (1990). The Irish Flute Players Handbook. Cork, Breac Publications.

Hurley, B. (2005, April 2006). "An Interview with Eamonn Cotter." from <http://www.firescribble.net/flute/cotter.html>.

Keegan, N. S. (1992). The Words of Traditional Flute Style. Music Department. Cork, University College Cork. **Masters Degree**.

Kohonen, T. (2001). Self-Organizing Maps, Springer Verlag.

Krassen, M. (1975). O' Neil's Music of Ireland, Waltons.

Larson, G. (2003). The Essential Guide to Irish Flute and Tin Whistle, Mel Bay Publications, Inc.

León, P. J. P. d. and J. M. Iñesta (2004). "Musical style classification from symbolic data: A two-styles case study." Lecture Notes in Computer Science **2771**: pp. 166-177.

Mandel, M. I., G. E. Poliner, et al. (2005). "Support vector machine active learning for music retrieval." ACM Multimedia Systems Journal.

Meyer, L. B. (1989). Style and Music. Theory, History and Ideology. Philadelphia, University of Pensylvania Press.

Quinlan, J. R. (1986). Induction of decision trees.

Repp, B. H. (1992). "Diversity and Commonality in Music Performance: an Analysis of Timing Microstructure in Schumann’s ‘Traumerei’." Journal of the Acoustical Society of America(104).

Roni. (2005, March 2005.). "The Amazing SlowDowner." from <http://www.ronimusic.com/>.

Salton, G. and C. Buckley (1988). "Term-weighting approaches in automatic text retrieval." Information Processing and Management **5**(24): 513-523.

Smith, E. and D. Medin (1981). Categories and Concepts, Harvard University Press.

Tansey, S. (1975). King of the Concert Flute.

Tansey, S. (1999). The Bardic Apostles of Innisfree, Tanbar Publications.

Tansey, S. (2006). Interview Notes. The Cobblestone Pub, Dublin.

Vallely, F. (1999). The Companion to Irish Traditional Music, New York University Press.

Vallely, F. (2004). Flute Routes to 21st Century Ireland, National University of Ireland. **Doctor of Philosophy**.

Wallis, G. and S. Wilson (2001). The Rough Guide to Irish Music. London, Rough Guides.

Widmer, G., S. Dixon, et al. (2005). "From Sound to `Sense' via Feature Extraction and Machine Learning: Deriving High-Level Descriptors for Characterising Music." Sound to Sense:Sense to Sound: A State-of-the-Art.

Widmer, G. and W. Goebl (2004). "Computational Models of Expressive Music Performance: The State of the Art." Journal of New Music Research **33**(3): 203–216.

Williamson, A., S. Thompson, et al. (2006). Creativity, originality and value in music performance. Musical Creativity Multidisciplinary Research in Theory and Practice. I. Deliége and G. A. Wiggins, Psychology Press.

Wright, M. and E. Berdahl (2005). "Machine Learning of Expressive Microtiming in Brazilian and Reggae Drumming." CS229.