Dublin Institute of Technology
ARROW@DIT

Conference papers

School of Computing

2009-01-01

Compensating for expressiveness in queries to a content based music information retrieval system

Bryan Duggan
Dublin Institute of Technology, bryan.duggan@comp.dit.ie

Brendan O'Shea
Dublin Institute of Technology, brendan.oshea@comp.dit.ie

Mikel Gainza
Dublin Institute of Technology, mikel.gainza@dit.ie

Padraig Cunningham
University College Dublin

Follow this and additional works at: http://arrow.dit.ie/scschcomcon

Part of the Computer Engineering Commons

Recommended Citation

Duggan, Bryan; O'Shea, Brendan; Gainza, Mikel; and Cunningham, Padraig, "Compensating for expressiveness in queries to a content based music information retrieval system" (2009). *Conference papers.* Paper 45.
http://arrow.dit.ie/scschcomcon/45

This Conference Paper is brought to you for free and open access by the School of Computing at ARROW@DIT. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@DIT. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie.

COMPENSATING FOR EXPRESSIVENESS IN QUERIES TO A CONTENT BASED MUSIC INFORMATION RETRIEVAL SYSTEM

Bryan Duggan, Brendan O' Shea

DIT School of Computing
Kevin St
Dublin 8, Ireland
{bryan.duggan,
brendan.oshea}@comp.dit.ie

Mikel Gainza

Audio Research Group
DIT Kevin St
Dublin 8, Ireland

mikel.gainza@.dit.ie

Pádraig Cunningham

School of Informatics and
Computer Science
UCD
Dublin, Ireland
padraig.cunningham@ucd.ie

ABSTRACT

MATT2 is a content based music information retrieval system adapted to the characteristics of traditional Irish dance music. MATT2 compensates for expressive artefacts commonly employed by traditional musicians. Specifically these are ornamentation, "the long note", reversing and phrasing. In this paper we describe the main components of MATT2 and present an experiment where we demonstrate that using this higher level knowledge of melodic similarity in traditional Irish dance music results in a significant improvement in annotation accuracy over standard approaches from the MIR literature.

1. INTRODUCTION

In this paper we present the results of an experiment where the effect of compensating for the presence of expressive artefacts in audio queries to a CBMIR (Content Based Music Information Retrieval) system is investigated. When an Irish traditional musician plays a tune, it is never played exactly as written. Specifically, our CBMIR system MATT2 (Machine Annotation of Traditional Tunes) compensates for four expressive artefacts: ornamentation, "the long note", reversing and phrasing [1].

Ornamentation in Irish traditional music is played on the beat, and alters the onset of the notes [2]. 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. An ornament can sometimes be replaced by *"the long note"*, whereby the note is simply held for the duration as an alternative [3].

Reversing is a technique originating in the Donegal fiddle tradition that is also popular amongst players of other traditional instruments. Reversing occurs when musician transposes down a part of a melody by one octave. This commonly occurs in the B part of a tune which is usually played in the high register of an instrument.

Phrasing in concert flute and tin-whistle music (two of the most popular of traditional instruments) is easily identified as the timings in a performance of a tune where a musician takes a breath. Taking a breath usually means leaving out one or more notes from the score, in a performance. Keegan [4] in his interviews establishes that phrasing (and in particular the length of phrases) is a strong indicator of a particular regional and individual style.

Section 2 presents an overview of MATT2 with an emphasis on how MATT2 compensates for the expressive artefacts described above. Section 3 introduces our experiment, which compares this approach with two alternatives common in the MIR (Music Information Retrieval) literature that make no accommodation for expressiveness. Following this, section 4 presents the results and statistical significance of this work using McNemar's test [5]. Section 5 presents conclusions.

2. MACHINE ANNOTATION OF TRADITIONAL TUNES (MATT2)

MATT2 works on mono digital audio files in the WAV (Waveform Audio Format) format recorded at 44.1KHz and was developed in Java[1]. A high level diagram of the sub-systems from MATT2 are presented in Figure 1.

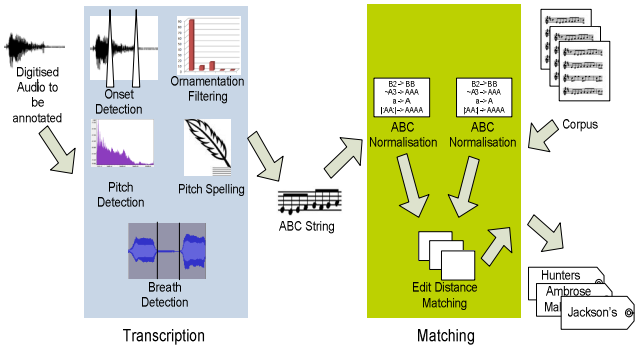


Figure 1: High level diagram of MATT2

The audio file to be annotated is first segmented into candidate note onsets and offsets using an onset detection function adapted from Gainza [6]. Onsets and offsets are considered by the system to be concurrent as traditional

[1] MATT2 and all test audio used in the experiments described in this paper can be downloaded from `http://www.comp.dit.ie/bduggan/music`

music is typically played legato and so a candidate note is considered to be a segment bounded by two adjacent onsets.

To establish the perceived pitch of each note, the *fundamental frequency (F0)* of the note is derived. MATT2 makes use of a frequency domain harmonic energy-based pitch detection algorithm based on Klapuri's multi-pitch estimator [7], adapted for monophonic transcription. The nearest match for the detected F0 frequency is the assigned the pitch spelling in ABC notation [8].

MATT2 incorporates an energy based breath detector subsystem to transcribe a breath in the signal. A breath is marked as a "z" in the transcription if average amplitude of a candidate note is less than a 10% threshold of the average amplitude over the entire signal.

To compensate for the playing of ornamentation notes, MATT2 makes use of an algorithm known as Ornamentation Filtering. This algorithm works on windowed segments of the input audio and is adaptive to tempo deviation, common in traditional music. The basic principle behind this algorithm is that clustering of notes based on relative note durations is used to identify ornamentation notes. These notes are removed from the transcription and the subsequent notes are lengthened by the duration of the ornamentation notes. Similarly, long notes are split into multiple quaver length 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 [2,4,9]. An earlier version of this algorithm is described in [1] and will be presented in the first author's forthcoming PhD thesis.

The corpus used in the experiment described in section 3 is Norbeck's reel and jig collection, which contains over 1500 reels and jigs, as well as variations of the tunes [10]. Before edit distance matching against the corpus is carried out, both transcribed string and strings from the corpus are normalised. This step is necessary as ABC notation supports features such as repeated sections, which need to be expanded so that they can be correctly matched against transcribed phrases. Normalisation involves the following 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. Triplets (melodically significant) are not removed. Secondly, all notes of duration greater than that of a quaver are expanded to be multiple instances of a quaver.

Thirdly, repeated sections are expanded and bar divisions are removed. ABC notation supports several notations for different types of repeated phrases [8]. This means for example, that if the transcribed query 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. In this way the system compensates for reversing as described in section 2. Figure 2 shows examples of each stage in the ABC normalisation process.

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 2: Normalisation stages for the A part of the tune "Come West Along the Road"

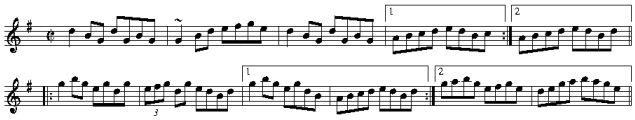
The minimum edit distance for each string from the corpus is then calculated using a cost of one for insertions, deletions and substitutions. In order to take a breath, a musician must leave out a note. Therefore, the edit distance cost function is adapted so that breath marks in ABC notation ("z") are allowed to match any character. In this way, phrasing is accommodated. A variation of the classic edit distance algorithm described by Navarro & Raffinot [11] 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. The system returns matching melodies from the corpus in order of lowest distance from the transcribed query.

3. EVALUATION

In order to evaluate the effect of expressiveness compensation in MATT2, the system was compared with two alternative approaches common in the MIR literature that do not make any accommodation for expressive performance. More than thirty musicians made test recordings for this experiment. Field recordings were made in imperfect conditions such as a kitchen in a house, a school room, various concerts in public halls and various public sessions, and contain ambient noise such as chairs moving, doors opening, foot taps and crowd noises. The recordings were edited so that the audio being tested contained fifty whole tunes (WT) and fifty short excerpts (E) from tunes. Deliberately challenging audio was

included, including degraded archive recordings, flute duets, flute and fiddle duets, fiddle solos, sessions with ensembles of up to ten musicians and ensemble playing in unusual keys with background noise. The following three scenarios are evaluated:

MC-ED: An edit distance matching algorithm based on melodic contours. This approach is common in the literature and is similar to the approaches employed by [12-15]. To perform this experiment, the corpus was first converted to MIDI format using the open source ABC2MIDI program from the ABC Music Project [16]. ABC2MIDI creates a MIDI rendering of a transcription in ABC format. Significantly, ABC2MIDI creates MIDI versions of any ornaments included in the transcription. The sequence of MIDI note numbers was extracted from each file and an algorithm was developed to convert this sequence to a melodic contour of "U", "D" and "S" characters, which denote up, down and same pitch respectively. Some examples of the output of the algorithm are given in Figure 3.



```
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||
74,71,67,74,67,71,67,69,67,71,74,76,78,
79,76,74,71,67,74,67,71,67,69,71,72,74,
76,74,71,72,74,71,67,74,67,71,67,69,67,
71,74,76,78,79,76,74,71,67,74,67,71,67,
69,71,72,74,76,74,71,74,79,83,79,76,79,
74,79,76,78,79,74,79,76,74,71,74,79,83,
79,76,79,74,71,69,71,72,74,76,74,71,74,
79,83,79,76,79,74,79,76,78,79,74,79,76,
74,71,74,79,81,83,79,76,78,79,76,74,76,
79,81,83,81,79,76
DDUDUDUDUUUUUDDDDUDUDUUUUUDDUUDDUDUDUDU
UUUUDDDDUDUDUUUUUDDUUUDDUDUDUUDUDDDUUUD
DUDDDUUUUDDUUUDDUDUDUUDUDDDUUUUDDUUDDUU
UUDDD
```

Figure 3: Various representations of the tune "Come West Along the Road"

The transcription system was adapted so that instead of quantising to the nearest playable note, the detected pitches were spelled as the closest MIDI note numbers. MIDI notes are quantised to the nearest semitone. From this sequence of MIDI note numbers, the melodic contour was again generated. Matching was performed using Navarro & Raffinot's [11] substring edit distance algorithm, which uses dynamic programming techniques.

TI-ED: A transposition invariant edit distance matching between corpus strings and transcribed queries was tested in the second scenario. For this experiment, the expressiveness compensation algorithms (Ornamentation Filtering, "long note" splitting, reversing, phrasing) described in section 2 were not employed. This was carried out to evaluate the specific effect of these algorithms. To perform this experiment, distances were calculated using Navarro & Raffinot's [11] substring edit distance algorithm between the MIDI note sequences for the query and MIDI note sequences derived from strings from the corpus as described above. Lemstrom's [17] transposition invariant edit distance cost function was employed to calculate distances.

MATT2: The complete system as described in section 2. For each experiment, the system annotated the test audio as described in section 2. In this experiment, the expressiveness accommodation algorithms of Ornamentation Filtering and ABC normalisation were employed.

For each of the three scenarios, the results were validated by a human expert who verified the accuracy of the annotations by proof listening to confirm that the retrieved scores were correct. Each test audio file was annotated with the meta-data from the corpus string with the minimum distance. In this way, the experiment only considered true positives *TP* and false positives *FP*. If there are *T* audio files to be annotated, then scores for *accuracy* and *error* are calculated as per (1).

$$accuracy = \frac{TP}{T} \qquad error = \frac{FP}{T} \tag{1}$$

4. RESULTS

Table 1 presents the accuracy and error scores for MC-ED, TI-ED and MATT2 for the 50 whole tunes (WT), the 50 excerpts (E) and the combined scores (C).

	MC-ED			TI-ED			MATT2		
	WT	**E**	**C**	**WT**	**E**	**C**	**WT**	**E**	**C**
TP	10	1	11	28	19	47	47	46	93
FP	40	49	89	22	31	53	3	4	7
Tot	50	50	100	50	50	100	50	50	100
Acc	0.20	0.02	.11	0.56	0.38	.47	0.94	0.92	.93
Err	0.80	0.98	.89	0.44	0.62	.53	0.06	0.08	.07
Tot	1	1	1	1	1	1	1	1	1

Table 1: Results for MC-ED, TI-ED and MATT2

MC-ED gives very poor accuracy and a high error rate for both WT and E. TI-ED is able to successfully annotate about half the whole tunes and less than half of the excerpts. MATT2 gives greater than 90% accuracy for both WT and E. When the results are combined, it can be seen that MC-ED gives 11% accuracy, TI-ED gives 47%

accuracy and MATT2 gives 93% accuracy. To establish the statistical significance of the results given in Table 1, contingency tables [5] are presented in Table 2.

A

53	36
0	11

B

7	82
0	11

C

7	46
0	47

Table 2: Contingency tables for MC-ED with TI-ED (A) MC-ED with MATT2 (B) and TI-ED with MATT2 (C)

The X^2 value for MC-ED and MATT2 (Table 2, B), is 80.01. The X^2 value for TI-ED and MATT2 (Table 2, C) is 44.02. That both of these values are above 3.841459 indicates that there is a statistical significant improvement in the performance of MATT2 compared to MC-ED and TI-ED.

5. CONCLUSIONS & FUTURE WORK

From the results of the experiment, it can be concluded that MATT2 substantially improves on pitch contour representations of music strings applied to MIR for traditional Irish music. Pitch contour representations give very poor accuracy when queries and the corpus contain ornamentation. Further, to the authors knowledge, MATT2 represents a unique attempt to adapt CBMIR to the specific characteristics of traditional Irish dance music. In comparing MATT2 with a SEMEX like approach [17], it is evident that the proposed system substantially improves accuracy over systems that have no specific accommodation for expressiveness. Results given demonstrate that this approach contributes to a significant improvement in accuracy when compared with an approach which uses a transposition invariant cost function, but does not possess any higher level knowledge about the music being analysed. In successfully testing MATT2 on audio acquired from real world sources it can be further concluded that the approaches outlined in this paper are robust to variations in musician, style and instrument. It is hoped that the work presented in this paper 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.

6. REFERENCES

[1] B. Duggan, B. O'Shea, and P. Cunningham, "A System for Automatically Annotating Traditional Irish Music Field Recordings," *Sixth International Workshop on Content-Based Multimedia Indexing, Queen Mary University of London, UK*, Jun. 2008.

[2] G. Larsen, *The Essential Guide to Irish Flute and Tin Whistle*, Mel Bay Publications, Inc., 2003.

[3] B. Breathnach, "Ceol Rince na hÉireann Cuid IV [Dance Music of Ireland] Vol IV," 1996.

[4] N. Keegan, *The Words of Traditional Flute Style*, MPhil Thesis, University College Cork, Music Department, 1992.

[5] T. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Computation*, vol. 10, 1998, pp. 1895-1923.

[6] M. Gainza, *Music Transcription within Irish Traditional Music*, PhD Thesis, Dublin Institute of Technology, Faculty of Engineering, 2006.

[7] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, 2003, pp. 804-816.

[8] S. Mansfield, Available at: http://www.lesession.co.uk/abc/abc_notation.htm [Accessed 14 May 2009]

[9] F. Vallely, *Timbre: The Wooden Flute Tutor*, Dublin, Ireland: Walton Manufacturing Company Ltd, 1986.

[10] H. Norbeck, Available at: http://www.norbeck.nu/abc/index.html [Accessed 14 May 2009]

[11] G. Navarro and M. Raffinot, *Flexible Pattern Matching in Strings: Practical On-Line Search Algorithms for Texts and Biological Sequences*, Cambridge University Press, 2002.

[12] A. Ghias, J. Logan, D. Chamberlin, and B. Smith, "Query by humming: musical information retrieval in an audio database," *Proceedings of the third ACM international conference on Multimedia*, 1995, pp. 231-236.

[13] L. Lu, H. You, and H. Zhang, "A new approach to query by humming in music retrieval," *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2001.

[14] S. Rho and E. Hwang, "FMF (Fast Melody Finder): AWeb-Based Music Retrieval System," *Computer Music Modeling and Retrieval: International Symposium, CMMR 2003, Montpellier, France, May 26-27, 2003: Revised Papers*, 2004.

[15] L. Prechelt and R. Typke, "An interface for melody input," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 8, 2001, pp. 133-149.

[16] S. Shlien, Available at: http://abc.sourceforge.net/abcMIDI/ [Accessed 14 May 2009]

[17] K. Lemstrom and E. Ukkonen, "Including interval encoding into edit distance based music comparison and retrieval," *Proceedings of the AISB 2000 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science , Birmingham*, 2000, pp. 53-60.