2010-01-01

Locating Tune Changes and Providing a Semantic Labelling of Sets of Irish Traditional Tunes

Cillian Kelly
Dublin Institute of Technology, cillian.kelly@dit.ie

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

David Dorran
Dublin Institute of Technology, david.dorran@dit.ie

Eugene Coyle
Dublin Institute of Technology, Eugene.Coyle@dit.ie

LOCATING TUNE CHANGES AND PROVIDING A SEMANTIC LABELLING OF SETS OF IRISH TRADITIONAL TUNES

Cillian Kelly, Mikel Gainza, David Dorran and Eugene Coyle
Audio Research Group
DIT Kevin St.
Dublin 8
Ireland
cillian.kelly@dit.ie

ABSTRACT

An approach is presented which provides the tune change locations within a set of Irish Traditional tunes. Also provided are semantic labels for each part of each tune within the set. A set in Irish Traditional music is a number of individual tunes played segue. Each of the tunes in the set are made up of structural segments called parts. Musical variation is a prominent characteristic of this genre. However, a certain set of notes known as 'set accented tones' are considered impervious to musical variation. Chroma information is extracted at 'set accented tone' locations within the music. The resulting chroma vectors are grouped to represent the parts of the music. The parts are then compared with one another to form a part similarity matrix. Unit kernels which represent the possible structures of an Irish Traditional tune are matched with the part similarity matrix to determine the tune change locations and semantic part labels.

1. INTRODUCTION

The approach presented here is specific to Irish Traditional Music. This music type consists of structural segments called 'tunes' which are concatenated to form 'sets'. The tunes are themselves made up of shorter structural segments called 'parts'. The structure of Irish Traditional Music is illustrated in Figure 1. Within this music type performers are encouraged to introduce musical variation. Parts which are notated as equivalent are aurally different due to this musical variation. The approach presented in this paper has two aims. The first is to determine the locations where tune changes occur within 'sets' of Irish Traditional tunes. The second aim is to assign a semantic label to each of the parts of each of the tunes within the music.

The information provided by a structural segmentation can be used for audio browsing. Instead of browsing through the music manually, using the structural segmentation information the user can browse directly to the part of interest within the music. Looping is a further application of the information provided by structural segmentation. Once a user has browsed to the required part within the music, the part can be looped to facilitate repeated playback of a certain segment. The structural segmentation information can provide exact loop points so that the start and end

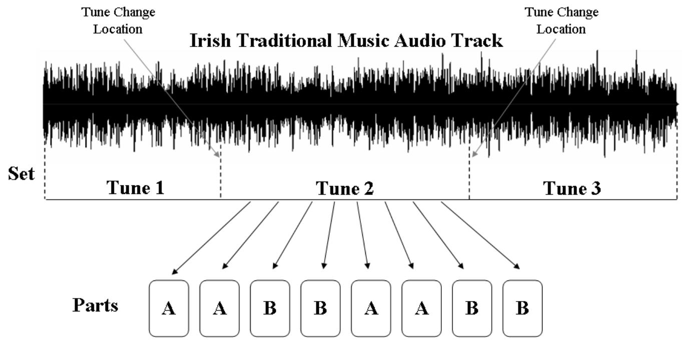

Figure 1. A representation of the structure present within a piece of Irish Traditional music. There are two distinct hierarchical levels of segmentation. The piece of music consists of segments called tunes, and each tune consists of further segments called parts.

of the selected loop will align rhythmically. This promotes aural learning which is common practice for Irish Traditional musicians. Structural segmentation information can also be used to create an audio thumbnail. For popular music, an audio thumbnail is the most repeated segment, often considered to be the chorus. For Irish Traditional Music however there is no chorus and no segment which repeats more often than others. Using the structural segmentation information a reduced form of the music can be created by discarding repeated sections.

Detailed in this paper is an approach which aims to extract tune change locations within sets of Irish Traditional music. The approach also provides a semantic labelling for each part of the tunes within each set. A set of notes known as 'set accented tones' [17] are utilised which are considered impervious to any musical variation. These notes are extracted and are used to represent the music. Using the harmonic information of these notes, sets are segmented into separate tunes, semantic labels are provided for each part of each resulting tune using a kernel matching technique.

The remainder of this paper is divided as follows: Section 2 details previous relevant approaches toward the structural segmentation of music. In Section 3 an overview of the structure of Irish Traditional Music is provided. The proposed approach toward locating tune changes and providing a semantic labelling of sets of Irish Traditional tunes is detailed in Section 4. Section 5 provides the results of testing the approach on a database of Irish Traditional tunes. The presented approach is compared directly with a previous approach which also attempts to locate the tune change locations within sets of Irish Traditional tunes. Finally, in Section 6 conclusions based on the results are provided.

2. LITERATURE REVIEW

Approaches which provide a structural segmentation of music aim to search for similarities within the audio signal. The signal is divided into audio frames and certain audio features are extracted for each of the resulting audio frames. Low level audio features such as zero crossings or spectral centroid are extracted and are combined to produce an aggregated description of the audio signal such as in [8, 11, 16, 21]. Mel Frequency Cepstral Co-efficients (*MFCCs*) are a further example of a low level audio feature which is commonly used for approaches which attempt to extract structure from music. In [13] *MFCCs* are extracted for each audio frame and the resulting frames are clustered in order to locate the repeated phrase or chorus within a 'rock' or 'pop' song. Certain heuristics are then used to choose a key phrase which corresponds to the chorus. These low level audio features are indicators of the timbre and loudness of the music. For Irish Traditional music, the timbre and loudness often remain constant throughout an entire musical piece therefore using these features is not suitable for this musical genre.

The audio features used for structural segmentation approaches may also be of a higher level of abstraction, such as pitch [2, 14] or chroma [1, 7] (See Section 4.3). Sections of the audio which share similar audio feature values are grouped together. The resulting groups of frames indicate the overall structure of the music. Extracting pitch or chroma values for every audio frame of an Irish Traditional music piece would include any musical variation present within the piece. This increases the difficulty of determining which structural segments are similar.

In [5], an approach is presented which segments audio using a measure of audio novelty. A Short Time Fourier Transform (*STFT*) is applied to the signal and each resulting frame is compared with every other frame to create an audio similarity matrix. The method of comparison used in [5] is the cosine distance measure. Points of significant musical change are determined by using kernel correlation. In [5], a checkerboard unit kernel is correlated along the diagonal of the resulting self-similarity matrix. Locations which result in a high correlation value are considered to be points of significant musical change which themselves are considered possible structural segmentation boundaries.

Structural segmentation of Irish Traditional music has attracted the attention of researchers. There have been a number of approaches which have attempted to structurally segment this music type [3, 4, 9, 10].

In [9] an approach is presented which aims to segment Irish Traditional tunes into their constituent parts and to provide a semantic labelling for the resulting parts. The 'set accented tones' which are considered impervious to variation are located within the music using a beat tracker. Pitch values are determined at these specific locations using a pitch detector. This results in a selective pitch contour. Melodic patterns are searched for amongst this pitch contour to determine the overall structure of the music. This approach was tested on a database of monophonic pieces of Irish Traditional music. The approach presented in [9] is extended further in [10] where chroma is calculated at 'set accented tone' locations rather than single pitch values. Following this, the chroma vectors are grouped according to heuristics specific to Irish Traditional Music. The resulting groups of chroma vectors correspond to the structural segments of the music and are compared using three different distance measures to determine which of the segments are similar. Extracting chroma rather than single pitch values at 'set accented tone' locations allows the approach in [10] to be applied to polyphonic music rather than only monophonic music as in [9].

In [4] an approach is presented which aims to provide the locations of tune changes (see Section 3) within a set of Irish Traditional tunes. This approach relies on a pre-existing database of transcribed Irish Traditional tunes. The music is transcribed using a pitch detection algorithm and is converted into a format consistent with the ABC music notation language [20]. Sections of the audio are compared with tunes contained within a database of ABC notated tunes using the edit distance algorithm [12]. Once the identity of the tune contained within the section has been determined the version of the tune from the database is compared with every possible section of the transcribed music again using the edit distance. This allows the algorithm to determine where the end of the current tune is located within the music. This process repeats for subsequent tunes within the set until all tunes have been processed.

The approaches presented in both [9] and [10] were tested on a database of pieces of Irish Traditional music containing one tune only. There is no attempt made to structurally segment sets of Irish Traditional tunes. The approach presented in [4] specifically addresses the problem of segmenting a set of Irish Traditional tunes by providing the locations of tune changes within the music. However in [4] the requirement of a pre-existing database of Irish Traditional tunes is a notable limitation. If a tune within a set is not contained within the pre-existing database of tunes, the segmentation of that set will not be successful. The approach presented in Section 4 attempts to provide a semantic labelling of a piece of Irish Traditional music as in [9]. However, unlike [9] the approach presented in Section 4 aims to provide this semantic labelling for sets of tunes rather than single tunes. Section 4 also details a method to locate each tune change within a set of Irish Traditional tunes as in [4]. A method using unit kernels is detailed which overcomes the requirement in [4] of a pre-existing database of transcribed tunes.

3. IRISH TRADITIONAL MUSIC

Irish Traditional Music is comprised of short musical pieces called tunes. Each tune is made up of two or more 'parts' which are notated using upper case letters as can be seen in Figure 1. Although each tune is quite short (a typical two part tune consists of sixteen bars), the parts are repeated to extend the tune and the tune itself can also be repeated in its entirety.

In both studio recordings and live performances of this music type often two or more tunes are concatenated into 'sets' to extend the music even further as shown in Figure 1. For example, a piece of music consisting of a two part tune followed by a three part tune may be played with a musical structure of *AABBAABB/AABBCCAABBCC*.

While two renditions of the same 'A' part may be notated identically, they are rarely performed identically. This is due to the large presence of musical variation inherent with this music type. Embellishments introduced by a musician will render two identical parts as being aurally different.

Despite the considerable presence of musical variation within this genre, for each tune there are a certain set of notes which are left unchanged by the musician. These notes are called the

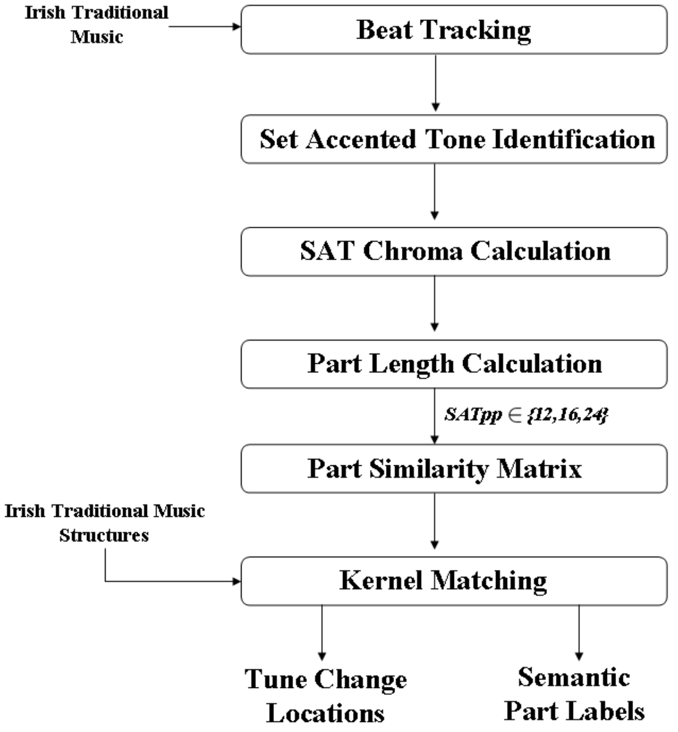

Figure 2. A block diagram of the proposed approach.

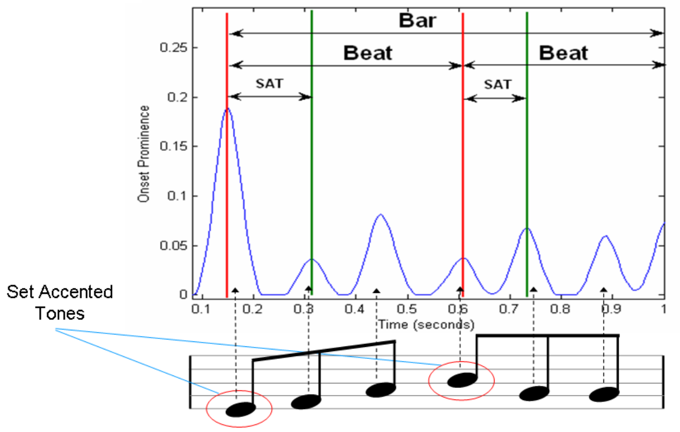

Figure 3. An onset detection function of one bar of Irish Traditional music. Each 'set accented tone' is located between the start of the beat and the next detected onset. For each 'set accented tone' a window is created between these two points of the onset detection function. Chroma is calculated for each resulting window.

'set accented tones' of the tune [9, 17]. In order to avoid the influence which musical variation has on determining which parts are equivalent, the 'set accented tones' are used to characterise the music for this approach.

4. PROPOSED APPROACH

4.1 Overview

An approach toward locating tune changes and providing a semantic labelling of sets of Irish Traditional tunes is detailed in this section. This approach is illustrated in the block diagram in Figure 2. The locations of the 'set accented tones' are determined using a beat tracker. Following this, chroma vectors are calculated at the resulting 'set accented tone' locations and are compared to create a part similarity matrix. Kernel matching is performed on the resulting matrix using unit kernels which represent the various musical structures present within this genre. The kernel matching technique presented here provides a solution to both determining the location of tune changes within a set and also to assigning a semantic label to each resulting part.

4.2 Beat Tracking and Set Accented Tone Identification

To extract chroma at 'set accented tone' locations within the music, the locations of the 'set accented tones' must be defined. Within Irish Traditional Music these notes are considered to be the first note of each beat. Therefore a beat tracker is employed to determine the location of each beat within the music. The beat tracker used for this approach is detailed in [6]. The beat tracker provides the location of each beat of the music along with an onset detection function which provides the locations of each note within the music. To encapsulate each 'set accented tone' a window is created extending from the start of each beat to the next detected onset as illustrated in Figure 3. This maximises the available harmonic information when determining chroma values at each 'set accented tone' location. Following the creation of each 'set accented tone' window, chroma information is extracted at each of these locations.

4.3 Chroma Calculation

To create the part similarity matrix detailed in Section 4.6 chroma must be calculated at each 'set accented tone' location. In [10] results showed that extracting chroma at 'set accented tone' locations provide better structural segmentation results than extracting a single pitch value. As such, for this approach chroma will also be used to represent each 'set accented tone'. Chroma is a spectral representation of music in which frequencies are mapped onto a set of 12 chroma values which correspond to the 12 notes of the equal tempered scale [18].

To calculate the chroma a Harmonic Pitch Class Profile (*HPCP*) approach is employed [19]. For each 'set accented tone' window (the section denoted as 'SAT' in Figure 3), a Short Time Fourier Transform is applied with a frame length of 2048 samples. The local maxima contained within each of the resulting STFT frames are identified using a peak picking algorithm. Following this, the magnitudes of each frequency at each resulting peak location are added to the appropriate chroma bin according to the note of the musical scale to which the frequency most closely corresponds. Only frequencies between 130Hz and 3140Hz are considered for this approach as 130Hz is the frequency of the lowest note on a banjo which is the lowest note likely to be present within an Irish Traditional tune and 3140Hz is the third harmonic of the highest note of a standard tin whistle, the highest note likely to be present within an Irish Traditional tune. This gives an appropriately rich description of the frequency content of a given audio frame. This results in a chroma vector of twelve elements each containing the amount of each note which was present in the given 'set accented tone' window.

4.4 Part Length Calculation

Following chroma calculation at each 'set accented tone' location, it is necessary to determine how many 'set accented tones' per part there are in the piece of music. This is required in order to determine the correct groups of chroma vectors to use when creating the part similarity matrix in Section 4.5. According to the Irish Traditional Music heuristics detailed in [9] there can only be 12, 16 or 24 'set accented tones' per part (*SATpp*) in an Irish Traditional tune. Consequently, each of these three conditions are tested and a confidence score is calculated for each possible *SATpp* value. Following this, the chroma vectors representing the 'set accented tones' are divided into groups according to the *SATpp* value currently being tested. For example, if the current *SATpp* value is equal to 12, the chroma vectors are divided into groups of 12. The resulting groups of chroma vectors now represent potential parts of a tune. Following this, each potential part is compared with every other potential part and a confidence score is calculated based on these part comparisons. The *SATpp* value which results in the highest confidence score is the value used when creating the part similarity matrix in Section 4.5.

Individual chroma vectors are compared using the Euclidean Distance formula given in Equation (1).

$$D(v_1, v_2) = \sqrt{(\sum_{i=1}^{12}(v_1(i) - v_2(i))^2)} \qquad (1)$$

where v_1 and v_2 are the two chroma vectors being compared.

Entire parts are compared with one another using Equation (2). The resulting value *S* is low if the two parts being compared are similar, therefore *S* is a measure of the dis-similarity between two parts.

$$S = \frac{\sum_{n=0}^{N-1} D(v_1(n), v_2(n))}{N} \qquad (2)$$

where *N* is equal to the *SATpp* value currently being tested.

Finally, a confidence value *C* for the *SATpp* value being tested is calculated according to Equation (3).

$$C = 1/\frac{\sum_{m=0}^{M-1} S_{SATpp}}{M} \qquad (3)$$

where *M* is equal to the total number of part comparisons. The *SATpp* value which results in the greatest confidence value *C* is the value used to create the part similarity matrix in Section 4.5.

4.5 Part Similarity Matrix

As detailed in Section 4.4, once the number of 'set accented tones' per part has been determined, the part similarity matrix is created according to this *SATpp* value. The values of *S* (calculated in Section 4.4 using Equation (2)) associated with this particular *SATpp* value are used to create the part similarity matrix. These values of *S* indicate the similarity of each part of length *SATpp* with every other part of length *SATpp* within the music.

Positioning these values into a matrix results in a part similarity matrix of size *P* by *P* where *P* is equal to the total number of parts within the music. An example of a part similarity matrix is shown in Figure 5. The part similarity matrix is used along with unit kernels to determine the structure of the music as described in Section 4.6.

4.6 Kernel Matching

The following section details how unit kernels are matched with the part similarity matrix created in Section 4.5. Firstly, the kernels that are used for matching are described, along with justifications for using these particular kernels. The process of how the unit kernels are matched with the part similarity matrix is then detailed.

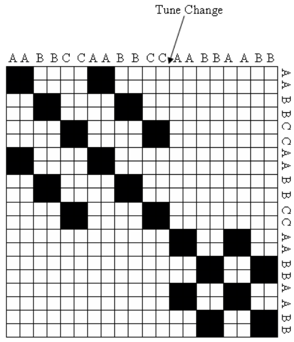

Figure 4. A unit kernel representing the structure of two tunes. The first tune represented has a structure of *AABBCCAABBCC* the second tune represented has a structure of *AABBAABB*. Black represents similar parts and white represents dis-similar parts.

Kernel matching relies on the availability of pre-existing unit kernels which each represent a specific musical structure. A unit kernel is a matrix consisting of ones and zeros which represent the pattern of a musical structure. A total of 24 unit kernels which represent a single tune are used here. Kernels representing the structure of more than a single tune are created by combining the kernels which represent a single tune. An example of a two-tune kernel is shown in Figure 4. There are 24 kernels used to represent the possible structures of one tune. As such combining each one-tune kernel with every other one-tune kernel results in 576 possible combinations for a two-tune kernel. The unit kernels represent the musical structures which are most common within Irish Traditional Music.

The kernels that are used limit the number of possible parts per tune to four. According to [15], tunes containing two, three and four parts make up 97% of the volume of tunes within this genre. Both one-tune kernels *and* two-tune kernels are utilised to give a total number of kernels of 600. The unit kernels are correlated with sections of the part similarity matrix as illustrated in Figure 5. The kernel which yields the highest matching value is the kernel which represents the most likely musical structure present within the given section of the part similarity matrix. The following steps describe the kernel matching technique:

1. At location *(i,i)* of the part similarity matrix, each *K x K* unit kernel is matched with a *K x K* section of the part similarity matrix using inner matrix multiplication. For the first iteration only, *i = 1*.

2. The kernel which results in the highest match value is deemed to represent the structure of that section of the part similarity matrix.

3. The value *i* is updated to be (*i + K*) where *K* is equal to the length of the kernel in step 2.

4. Steps 1-3 are repeated until the entire part similarity matrix has been processed.

The outcome of this kernel matching as outlined in Figure 2 is the location of each tune change within the set of Irish Traditional tunes along with a semantic label for each resulting part.

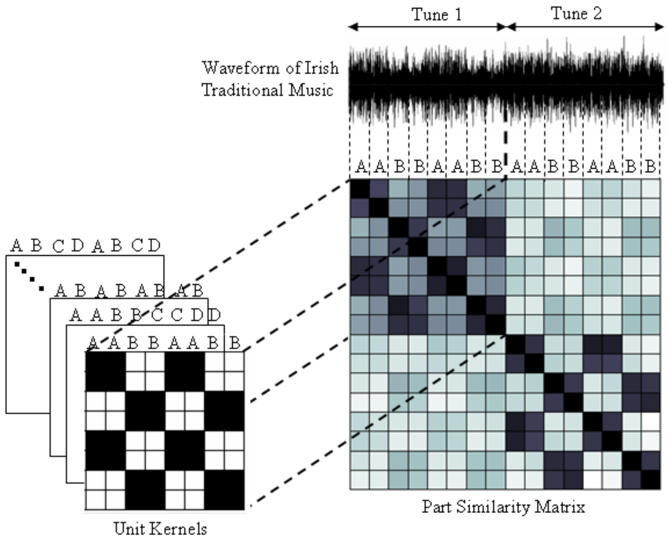

Figure 5. The section of a part similarity matrix with which a unit kernel is correlated. Each cell of the part similarity matrix corresponds to a part of an Irish Traditional tune. This section of the part similarity matrix is correlated with each unit kernel. The unit kernel which results in the highest matching value corresponds to the structure of this section of the part similarity matrix.

The semantic part labels are represented by the concatenation of the descriptions of the unit kernels which were matched to each tune. The locations of each tune change are determined by the beat locations of the start of each successfully matched unit kernel. Results for the approach which has been presented here are detailed in Section 5.

5. RESULTS

The evaluation of this approach was carried out on a hand annotated database of 30 sets of Irish Traditional music. These 30 musical pieces consist of 75 separate Irish traditional tunes with 34 tune changes and a total of 589 parts. The results of detecting the location of tune changes and determining labels for each structural segment were calculated separately. A tune change location was considered correct if the automatically detected tune change locations were within 1 second of the hand annotated tune change locations. Table 1 details the results of labelling the parts of the music and also details the results of locating tune changes within the music. In Table 1, *N* is equal to the number of annotations, *GP* is equal to Good Positives, *FP* is equal to False Positives and *FN* is equal to False Negatives. Results were calculated using three different measures, precision, recall and accuracy. These three measures are defined by Equations (4), (5) and (6).

$$Precision = \frac{GP}{GP + FP} \tag{4}$$

$$Recall = \frac{GP}{GP + FN} \tag{5}$$

$$Accuracy = \frac{N - FP - FN}{N} \tag{6}$$

The results show that the approach can label the parts contained in a piece of Irish Traditional Music with an accuracy of 86% along with a precision value of 90% and a recall value of 94%. Additionally, this approach can correctly identify the location of tune changes within a set of Irish Traditional tunes with an accuracy of 62% along with a precision value of 77% and a recall value of 80%.

	N	*GP*	*FP*	*FN*	*Prec*	*Rec*	*Acc*
Part Labels	589	479	51	31	90%	94%	86%
Tune Change Locations	34	24	7	6	77%	80%	62%

Table 1. Results of the part labelling and locating tune changes produced by the presented approach.

Tolerance	*Prec*	*Rec*
1 second	69.7%	64.79%
2 seconds	87%	81%

Table 2. Results of the approach presented in [3] and [4] toward locating tune changes within a set of Irish Traditional tunes.

The accuracy value is higher for labelling parts than for detecting tune changes. This is because even when a tune change is not accurately detected, the algorithm may still correctly identify subsequent parts contained within the music. This is due to many kernels having common part locations. The high recall values should be noted, this indicates that the algorithm detects most tune changes present in the music.

An approach is presented in [4] which also attempts to calculate the tune change locations within sets of Irish Traditional music. In [4] a tune change location is considered to be correct if the automatically generated tune change locations are within 2 seconds of the equivalent hand annotated tune change locations. The approach presented in [4] is detailed further in [3] where results are also provided for a tolerance of 1 second. The results of detecting tune change locations as detailed in [3] and [4] can be seen in Table 2 for a tolerance window of both 1 second and 2 seconds. The results detailed in Table 1 were obtained from testing on the same database used in [3].

When detecting tune changes within a set the approach detailed in [3] claims a precision value of 69.7% and a recall value of 64.79% for a tolerance of 1 second. The approach presented in Section 4 has improved these values by 7.3% and 15.21% respectively and also does not require the database of Irish Traditional tunes which is utilised in [3]. In [3], to correctly detect the tune changes within a set of Irish Traditional tunes, each tune in the set must also be in a pre-existing database of tunes.

6. CONCLUSIONS

This paper presented an approach toward locating tune changes and providing a semantic labelling of sets of Irish Traditional tunes. This music type consists of sets of tunes, the tunes themselves are made up of parts. This approach utilised certain notes within the music which remain constant despite the presence of musical variation. Chroma was extracted at these specific note locations and was compared to create a part similarity matrix. Unit kernels representing common structures present within Irish Traditional Music were then matched with sections of the part similarity matrix. The unit kernel which resulted in the highest match value corresponds to the structure of the music at the given location within the part similarity matrix.

Using chroma at the 'set accented tone' locations within the music significantly reduces the amount of data required to produce a structural segmentation. This reduced representation of the music also filters out musical variation which can affect the process of determining which parts are equivalent. The approach presented here was tested on a database of 30 sets of Irish Traditional tunes. The results of the approach presented here were compared with a similar approach toward detecting tune change locations within a set of Irish Traditional tunes by testing on the same database of Irish Traditional music. When a tolerance of 1 second between automatically detected tune changes and hand annotated tune changes is used, the approach presented here performs significantly better than a previous approach toward the same goal. For this approach, increasing the tolerance window will not result in an increase in performance as tune change location times are calculated using part locations and are not calculated on a scale which is sensitive to increments of less than the length of a part.

The approach presented in this paper relies on correctly calculating the amount of 'set accented tones' per part (*SATpp*). If this value is calculated incorrectly, the resulting part similarity matrix will not accurately reflect the parts which are present within the set of Irish Traditional tunes. Consequently, it would not be possible to identify the correct tune change locations or determine the correct semantic part labelling. This approach also relies on the accuracy of the beat tracker in order to correctly identify the 'set accented tone' locations.

Future work will aim to combine the approach presented here with the approach presented in [4] to identify tune change locations within a set. Firstly, the method detailed in [4] would be used to determine the tune change locations. If there are tunes present within a set that are not present within the pre-existing database used in [4] the tune change locations cannot be calculated using this method. In this case, the approach presented in Section 4 would be used as an alternative, as there is no pre-existing knowledge required of the particular tunes within the set for this approach.

7. REFERENCES

[1] Mark A. Bartsch and Gregory H. Wakefield. To catch a chorus: using chroma-based representations for audio thumbnailing. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, 2001.

[2] Roger B. Dannenberg and Ning Hu. *Discovering Musical Structure in Audio Recordings*. Lecture Notes in Computer Science. Springer Berlin, 2002.

[3] Bryan Duggan. *Machine Annotation of Traditional Irish Dance Music*. PhD thesis, Dublin Institute of Technology, 2009.

[4] Bryan Duggan, Brendan O'Shea, Mikel Gainza, and Padraig Cunningham. Machine annotation of sets of traditional irish dance tunes. In *International Conference on Music Information Retrieval*, Philadelphia, PA, USA, 2008.

[5] Jonathan Foote. Automatic audio segmentation using a measure of audio novelty. In *IEEE Intl Conf. on Multimedia and Expo*, New York, 2000.

[6] Mikel Gainza. On the use of a dynamic hybrid tempo detection model for beat tracking. In *IEEE International Conference on Multimedia and Expo*, Singapore, 2010.

[7] Masataka Goto. A chorus-section detecting method for musical audio signals. In *IEEE Conference on Acoustics, Speech, and Signal Processing*, Hong Kong, 2003.

[8] Min-Hong Jian, Chia Han Lin, and Arbee L.P. Chen. Perceptual analysis for music segmentation. In *Storage and Retrieval Methods and Applications for Multimedia*, San Jose, California, USA, 2003.

[9] Cillian Kelly, Mikel Gainza, David Dorran, and Eugene Coyle. Structural segmentation of music using set accented tones. In *124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, 2008.

[10] Cillian Kelly, Mikel Gainza, David Dorran, and Eugene Coyle. Structural segmentation of irish traditional music using chroma at set accented tone locations. In *127th Audio Engineering Society Convention*, New York, New York, U.S.A., 2009.

[11] S. Lefevre, B. Maillard, and N. Vincent. A two level classifier process for audio segmentation. In *16th International Conference on Pattern Recognition*, Washington DC, USA, 2002.

[12] Vladimir Iosifovich Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 1966.

[13] Beth Logan and Stephen Chu. Music summarization using key phrases. In *International Conference on Audio Speech and Signal Processing*, Istanbul, Turkey, 2000.

[14] Benoit Meudic. Musical pattern extraction: from reptition to musical structure. In *Computer Music Modelling and Retrieval*, Montpellier, France, 2003.

[15] Donncha Seán Ó'Maidín. *A Programmer's Environment for Music Analysis*. PhD thesis, University College Cork, 1995.

[16] Bee Suan Ong and Perfecto Herrera. Semantic segmentation of music audio contents. In *International Computer Music Conference*, Barcelona, Spain, 2005.

[17] Míchéal Ó'Súilleabháin. *Innovation and Tradition in the Music of Tommie Potts*. PhD thesis, Queen's University, 1987.

[18] Steffen Pauws. Musical key extraction from audio. In *International Conference on Music Information Retrievals*, Barcelona, Spain, 2004.

[19] Joan Serra, Emilia Gomez, Perfecto Herrera, and Xavier Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008.

[20] Chris Walshaw. Abc notation - an introduction, http://abcnotation.com/. Accessed March 2010.

[21] Yibin Zhang and Jie Zhou. Audio segmentation based on multi-scale audio classification. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Quebec, Canada, 2004.