2008-01-01

Structural Segmentation using Set Accented Tones

Cillian Kelly
Technological University Dublin, cillian.kelly@tudublin.ie

Mikel Gainza
Technological University Dublin, Mikel.Gainza@tudublin.ie

David Dorran
Technological University Dublin, david.dorran@tudublin.ie

See next page for additional authors

Audio Engineering Society

Convention Paper

Presented at the 124th Convention
2008 May 17–20 Amsterdam, The Netherlands

Structural Segmentation of Music Using Set Accented Tones

Cillian Kelly[1], Mikel Gainza[2], David Dorran[3], Eugene Coyle[4]

[1] Audio Research Group (Dublin Institute of Technology), Kevin St, Dublin 2, Ireland
cillian.kelly@dit.ie

[2] Audio Research Group (Dublin Institute of Technology), Kevin St, Dublin 2, Ireland
mikel.gainza@dit.ie

[3] Audio Research Group (Dublin Institute of Technology), Kevin St, Dublin 2, Ireland
david.dorran@dit.ie

[4] Audio Research Group (Dublin Institute of Technology), Kevin St, Dublin 2, Ireland
eugene.coyle@dit.ie

ABSTRACT

An approach which efficiently segments Irish Traditional Music into its constituent structural segments is presented. The complexity of the segmentation process is greatly increased due to melodic variation existent within this music type. In order to deal with these variations, a novel method using 'set accented tones' is introduced. The premise is that these tones are less susceptible to variation than all other tones. Thus, the location of the accented tones is estimated and pitch information is extracted at these specific locations. Following this, a vector containing the pitch values is used to extract similar patterns using heuristics specific to Irish Traditional Music. The robustness of the approach is evaluated using a set of commercially available Irish Traditional recordings.

1. INTRODUCTION

This paper introduces a novel technique to segment audio into its constituent structural segments. To achieve a structural segmentation certain notes within the music known as 'set accented tones' are utilised. These notes are extracted from the audio and are used to represent the audio in its entirety. The 'set accented tones' also provide a solution to the problem of melodic variation between two structural segments that are

perceived by humans to be similar. Structural segments of the audio are located by searching for patterns within the 'set accented tones'.

Section 2 provides an overview of previous approaches that attempt to structurally segment audio and certain problems with these approaches are highlighted. In Section 3, musical theory specific to Irish Traditional Music is presented to increase understanding of the proposed approach. Section 4 describes the method used to extract the 'set accented tones' from the audio and also provides a description of the pattern recognition techniques used to locate the structural segments. The results of this approach are presented in Section 5. Finally, Section 6 provides conclusions and details of future work.

2. EXISTING APPROACHES

There are many existing approaches that attempt to provide a method to structurally segment audio. These approaches are quite diverse and results vary depending on the method used. Using a measure of self-similarity is a method of segmentation employed in [1] [2] [3] [4] and [5]. In each case, the audio signal is divided into frames where each frame is compared against every other frame in the signal. The similarity between frames is calculated using a distance measure. This yields a two-dimensional array of similarity values. The array is typically visualised as colour values with brightness representing similarity and darkness representing dissimilarity. This visualisation is useful for detecting repeating patterns within the audio. The differences between each approach are the mid-level representations and the distance measures that are used.

In [1], the signal is represented by a spectrogram and the cosine distance measure is used. The method employed in [2] differs from [1] in that it calculates the beats of the audio and extracts musical information between each beat of the audio. The features extracted are rhythm, melodic contour and pitch intervals. These three features are used as the mid-level representations of the signal. Individual similarity matrices are computed for each feature. Following this, the three resulting matrices are compared with one another to discern similarities common to all three features. The structure of the music is determined by these common areas of similarity. A similar approach to [2] is presented in [3]. The musical features used here are rhythm, timbre and pitch. As in [2], these features are subjected separately to self-similarity analysis. The 'L2 norm' is the distance measure used. It is shown that using the timbre feature provides the most accurate results because humans tend to segment by timbre more so than segmenting by rhythm or harmony. Self-similarity is also an approach employed in [4], where many features are used as mid-level representations. Information pertaining to spectral centroids, sub-band energy, zero crossings, spectral roll-off, RMS energy, spectral flux, spectral flatness, high-medium energy and low bass energy are extracted for each frame. The cosine distance measure is used to compute a similarity matrix for each one of these features. Kernel correlation is applied along the diagonal of the matrices to compute novelty measures. Segments are detected by finding local maxima from the novelty measures. The detected peaks are then combined to give boundary candidates that correspond to segment changes in the audio. In [5], audio segmentation using self similarity based on past and future frames is introduced. Self-similarity is computed for the past and future of each frame. Based on these similarities, a measure of frame novelty is computed. Frames that display a high novelty score indicate where musical changes are most likely to occur. These points of significant change in the audio are considered to be structural segment boundaries.

Clustering is a segmentation technique that is utilised in [6], [7], [8] and [9]. It is an unsupervised learning process of combining features which share common characteristics. In [6], a clustering method applied to a mid-level representation of the audio is used to extract the chorus of a song. A song is described using 'Mel-Cepstral Features', which model the human auditory perception. The initial clusters are provided by dividing the song in to arbitrary segments. The clusters remaining when the algorithm concludes correspond to the structural segments. In [9], musical phrases are extracted from the audio and clustered into 'sentences' using rhythmic features and 'melodic shape'. The melodic shape of a phrase is defined by its first note, last note and the average pitch value of the remaining notes. This information is used to extract the musical sentences, which are constructed by clustering the various phrases. A 'sentence' is considered to be a group of one or more musical phrases which repeat within the music. These 'sentences' constitute the structure of the audio.

A technique to extract the chorus of a song using a chroma representation is outlined in [10] and [11]. This representation consists of 12-dimensional feature vectors, where each of the 12 values represents a note on the musical scale. These vectors are compared with

vectors from other locations in the signal to determine similarity. Repeated sections are identified and the most repeated section is labelled as the chorus.

In [12], three low-level features are used in order to segment audio. A measure of 'roughness' pertaining to changes in timbre, 'periodicity pitch' which is simply the pitch of a particular sound described by its period and 'loudness', which equates to the power present in each audio frame. These three features are used amid heuristic methods to determine candidate segment boundaries within the audio.

There are a number of notable problems present within these techniques. In [2], the approach has been designed for use with MIDI only. The effectiveness of the approaches used in [4] and [5] are reduced when there are very smooth musical transitions between segments. In addition, for the approach used in [4], eight separate similarity matrices are required which would be computationally expensive. The approaches outlined in [6] and [8] are based on clustering similar sections together according to timbre. These approaches will not work well when applied to solo performances where the same timbre will be displayed throughout the audio. Performance of the segmentation method outlined in [10] degrades when there are significant changes present between two renditions of the same chorus, although a solution to this problem is proposed in [11].

The approach presented in this paper attempts to provide a structural segmentation of audio that overcomes the problems associated with efficiency, labeling and melodic variation. A novel technique to summarize audio is also presented.

3. IRISH TRADITIONAL MUSIC THEORY

The approach presented in this paper deals specifically with the structural segmentation of Irish Traditional Music. For this reason, a certain amount of knowledge of the structure of this music type is required to better understand the proposed method. An Irish Traditional tune is made up of sections referred to as 'parts'. Each 'part' consists of either four or eight bars. The majority of Irish tunes are made up of two or three 'parts'. Each 'part' is denoted by a capital letter. So the first 'part' will be denoted 'A', the second 'B' as can be seen in Figure 1. 'Parts' are customarily repeated and the tune itself is also repeated in its entirety a number of times. Accordingly, the form or structure 'AABBAABBAABB' is one interpretation of a two 'part' tune containing an 'A' and 'B' 'part'.

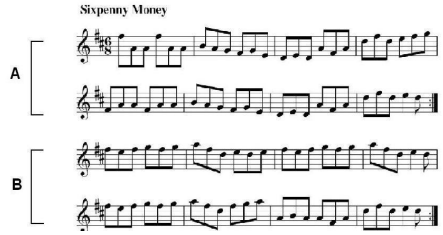

Figure 1. An Irish Traditional tune with two parts.

However, it is in the nature of this music type that any rendition of a tune is open to interpretation. So the structure of a particular tune could be different depending on the musician who is playing it. It is common that a considerable amount of melodic variation is present between two renditions of the same part within the same performance. Again, the nature of this variation is unique to each musician. There exists a set of notes that remain constant regardless of the amount of melodic variation present. These 'set accented tones' are referred to in [13] as "important accentuated points" and as the "stepping stones of a piece". It is also stated in [13] that these notes are "at the heart of the tune's identity, and that any extended interference with them is in the nature of contradiction of the tune itself". Thus, these particular notes are considered to be impervious to variation. Consequently the 'set accented tones' are representative of the tune due to their static nature. It is the existence of these notes that allows such a vastly reduced representation of the audio. These notes which are located directly after each beat are extracted from the audio and are used here to represent the audio track in its entirety. The position of the 'set accented tones' within an Irish Traditional tune is illustrated in Figure 2.

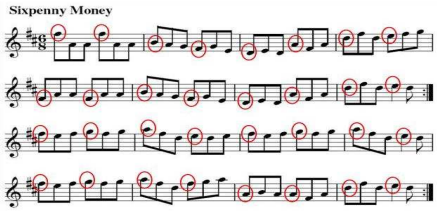

Figure 2. An Irish Traditional tune in 6/8 time with the 'set accented tones' highlighted.

There are a finite number of time signatures present within Irish Traditional Music. These time signatures,

along with the associated number of 'set accented tones' per part are outlined in Table 1.

Tune Type	Time Signature	Bars Per Part	Set Accented Tones Per Part
Reel Hornpipe March	4/4	8	16
Jig	6/8	8	16
Polka BarnDance	2/4	8	16
Slip Jig	9/8	4	12
Slide	12/8	4	12
Waltz	3/2	8	24
Mazurka	3/4	8	24

Table 1. An overview of the time signatures present within Irish Traditional Music and the bars per part and set accented tones per part associated with each time signature.

In general within Irish Traditional Music, each set accented tone is located on the beat of the music. A notable exception to this are tunes set to a time signature of 4/4. For a tune set in 4/4 there are 4 beats per bar. However it is shown in [13] that despite this fact there only exists two 'set accented tones' per bar. Consequently, for the case of tunes set in 4/4, only the alternating 1st and 3rd beats of the bar are considered during the 'set accented tone' extraction phase. The position of the 'set accented tones' within an Irish Traditional tune with a time signature of 4/4 is illustrated in Figure 3.

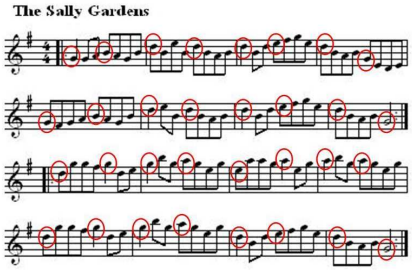

Figure 3. An Irish Traditional tune in 4/4 time with the 'set accented tones' highlighted.

The importance of providing structural segmentation for Irish Traditional Music lies within audio browsing. It is common practice within this genre to concatenate a number of different tunes to comprise a single track. Segmenting the audio will assist the user in navigating through different 'parts' of the track. In addition, the structural segments constitute a useful thumbnail for an audio track. By using segment locations, a more meaningful audio thumbnail can be generated. For instance, a tune that is played with the form 'AABBCCAABBCC' could be reduced to a thumbnail with the structure 'ABC' using the information provided by a structural segmentation. This is a more meaningful representation of an audio track than is currently used by online music stores, which generally play an arbitrary 30 second segment within a song. Segment locations are also useful in looping tools. For example, a musician can accompany a loop in order to support the tradition of aural learning in Irish Traditional Music.

4. PROPOSED APPROACH

4.1. Introduction

The approach proposed here uses 'set accented tones' to search for repeating patterns within an audio track. As can be seen in Figure 4, there are a number of steps required to extract each 'set accented tone' from the audio. These steps result in a vector of pitch values which is inspected for the presence of repeating patterns. The structure of the music is defined by these repeating patterns.

Most segmentation approaches start with a feature extraction step where the audio is divided into sections called frames. Feature vectors are calculated from these frames and the audio is described in terms of these features. The approach presented here employs a similar technique. A vector of pitches containing the 'set accented tones' is extracted from the audio. These tones provide the summarized representation of audio used throughout this approach. The 'set accented tones' vector is divided in to a number of possible 'part candidates'. These 'part candidates' are self-compared to discern which of the parts are similar. The parts are then labeled according to their similarities. The part labels along with the times at which each part begins within the audio results in a possible structural segmentation. Irish Traditional Music heuristics are then applied to determine whether the segmentation is plausible. A confidence measure is employed to determine the most likely structural segmentation.

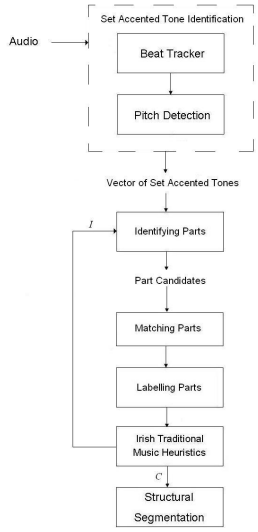

Figure 4. Flow Diagram of the structural segmentation approach using 'set accented tones'.

In Section 4.2 the 'set accented tones' are extracted from the audio using a beat tracker and a pitch detector. In Section 4.3 the 'set accented tones' are divided in to possible 'part candidates'. The lengths of these 'part candidates' are governed by Irish Traditional Music heuristics. In Section 4.4, similarity scores between each 'part candidate' are computed to discern which 'parts' should be considered to be similar. The similarity score between 'part candidates' must be over a certain threshold in order to be considered as a candidate. Details of how the algorithm deals with potential errors computed by the beat tracker and pitch detector by computing a number of iterations are also outlined. Section 4.5 gives details on the process of labeling each structural segment. Finally, Section 4.6 outlines the Irish Traditional Music heuristics that are used and gives details on the confidence measure that is used to determine the most likely structural segmentation result.

4.2. Set Accented Tone Identification

There are a number of steps required to extract the 'set accented tones' from the audio. Firstly, a beat tracker is applied to the audio track using a Short Time Fourier Transform. The beat tracker used here is outlined in [14]. This beat tracker calculates the time within the audio at which each beat location occurs. Following this, a pitch estimator is applied in the region of each 'set accented tone'. A window encapsulating the region of each 'set accented tone' location is calculated and pitches are estimated for the audio contained within each window. The windows used are illustrated in Figure 5. This figure shows the note onsets detected in a single bar of an Irish Traditional reel in 4/4 time. There are two 'set accented tones' denoted by an *S* present within the bar. The 'set accented tones' are located between the start of the first and third beats and the next detected onsets after those beats.

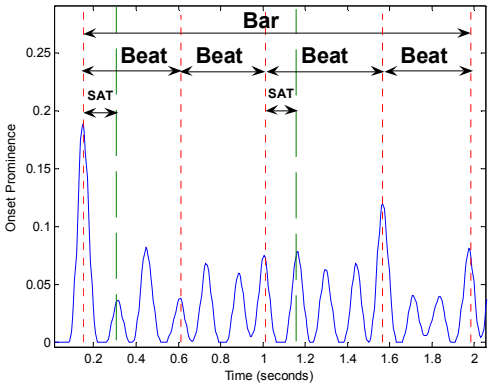

Figure 5. A graph showing the onsets detected for the first bar of an Irish traditional reel in 4/4 time. The dotted line corresponds to the beat locations; the dashed line corresponds to the next detected onset after a beat location. Each *SAT* denotes the location of a 'set accented tone'.

Within Irish Traditional Music, the first note immediately following each beat location corresponds to a 'set accented tone'. Consequently, each window length is equal to the distance between the start of a beat and the next onset. This maximises the window lengths in order to increase the accuracy of the pitch detection. Following this, pitch information is extracted for each window which results in a vector containing the 'set accented tones'. Each note contained within the vector is represented by a numerical value between one and twelve, with C being equal to one. Notes are octave independent. Subsequently, repeating patterns are searched for within this vector.

4.3. Identifying Parts

The pattern recognition technique uses information specific to Irish Traditional Music in order to generate different 'part' candidates. The information on meters within Irish Traditional Music outlined in Table 1 leads to an important hypothesis: there can only be twelve, sixteen, or twenty-four 'set accented tones' per part in any given Irish Traditional tune. Therefore, the algorithm undertakes a number of separate passes. The

first pass tests for the case where the tune could contain twelve 'set accented tones' per part, the second pass tests the case where the tune could contain sixteen 'set accented tones' per part and so on. Thus, for each pass, the 'set accented tone' vector is split into equal 'part' candidates. These 'parts' are then self-compared to determine which among them should be considered similar to each other.

A common problem encountered when using a beat tracker is that the estimated tempo of the music is occasionally doubled or halved. For this reason, the vector of 'set accented tones' is also split into part candidates with both the double and half of the number of 'set accented tones' per part of that particular pass. These parts are also examined for possible similarities.

4.4. Matching Parts

To determine which part candidates should be considered similar, the part candidates are compared melodically. The pitch values of each part candidate are compared against each other note for note as can be seen in Figure 6. A part similarity score P is stored which measures how similar two parts are to each other. P is calculated as follows:

$$P = \sum_{i=1}^{t} (S) \qquad (1)$$

where t is equal to the number of 'set accented tones' per part and S is equal to the similarity resulting from each note comparison. Each note comparison score S is computed as follows: if there is a direct melodic match between two notes, a score of 1 is assigned to that note. If the two notes differ by a semitone, a score of 0.6 is assigned and if the two notes differ by a full tone, a score of 0.2 is assigned. A score of 0 is assigned if the two notes differ by more than a full tone. These values were chosen arbitrarily through informal testing. If the part similarity score P is over a threshold T, those two parts are considered to be similar. The threshold T is defined as:

$$T = (M * 0.9) - I \qquad (2)$$

where M is the length of a part candidate and I represents the I^{th} iteration of the algorithm as explained below.

A possible segmentation result will be reached if the part similarity score P is greater than the threshold T for each part candidate. If any part matches no other part, that particular segmentation attempt is discarded.

Each pass of the algorithm also contains a number of separate iterations. If the first iteration provides no segmentation result, certain criteria are altered to allow for errors in both the pitch estimation and the beat tracker. Criteria are altered as follows: for errors with the pitch estimation, the threshold T that each part comparison must reach for the two part candidates to be considered similar is lowered by I as can be seen in Equation 2. Also, to combat errors with the beat tracker, the melodic comparison of notes within each part is expanded to also consider adjacent notes as illustrated in Figure 7 and Figure 8. It should be noted that subsequent iterations are only computed if the previous iteration produced no segmentation result.

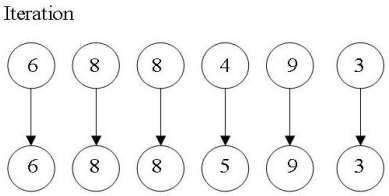

Figure 6. Notes from one part are compared directly with notes in equivalent positions from other parts with the numbers representing pitch values.

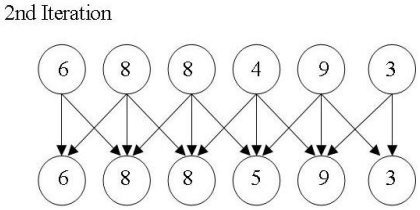

Figure 7. During the 2nd iteration, adjacent 'set accented tones' are also compared against to allow for possible spurious beats.

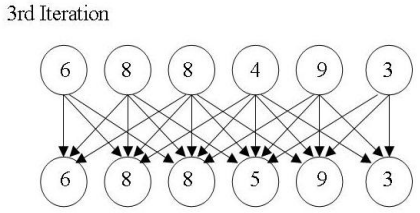

Figure 8. The comparisons during the 3rd iteration are expanded to include even more notes.

4.5. Labeling Parts

Labeling is carried out concurrently with the part matching. The first part is always labeled 'A'. The part matching from Section 4.4 is then carried out, any parts considered to be similar to this first 'A' part are also labeled 'A'. Once all parts have been compared against the first part, the algorithm backtracks to the earliest part which has not yet been assigned a label. This part is labeled 'B'. This 'B' part is then checked for similarity against all unlabelled parts. Following this, parts that are considered similar to the 'B' part are also labeled 'B'. This process of comparing and labeling unlabeled parts repeats until all parts have been assigned a label.

4.6. Validating Parts and Confidence Measure

4.6.1. Irish Traditional Music Heuristics

Possible segmentations are validated to determine if they are plausible within the heuristics of Irish Traditional Music. Certain rules are applied to ensure that the structural segmentation complies with these heuristics.

- A label may not occur three or more times consecutively.
- A label must appear within a distance of p*2 of another instance of that label, where p is equal to the number of unique labels within the tune.
- The amount of parts must be an even number.
- The last label must be the highest alphabetically.

If all of these criteria are met, the segmentation result can be considered plausible within the constraints of Irish Traditional Music. If all of the criteria are not met, the segmentation result is considered implausible and is discarded.

4.6.2. Confidence Measure

At this stage of the algorithm there will exist one or more possible segmentation results. A decision must be made as to which of the possible segmentations is most likely correct. A confidence measure is calculated and associated with each candidate segmentation resulting from each pass. A confidence score C is calculated as follows:

$$C_k = \left(\sum_{i=1}^{L_k} (P) \right) - (I * w) \qquad (3)$$

where k is the particular structural segmentation with which C is associated, L is the number of part candidates, P is the part similarity score for each part which was calculated in Section 4.4, I is the number of iterations the algorithm completed before it reached this particular segmentation and w is a weighting applied to the iteration number. The value w is required because the value I by itself will be too small a number to affect the result of the equation in any considerable way. Therefore it is multiplied by a larger value w to apply more significance to the number of iterations that have already occurred. After informal testing, the weighting value w was set to be equal to L * 2.

This global confidence measure is the defining value that will decide which segmentation will be considered correct upon completion of the algorithm. The segmentation with the highest confidence score after all passes are completed is returned as the most likely structural segmentation. The process of choosing the segmentation with the highest confidence score is illustrated in Figure 9.

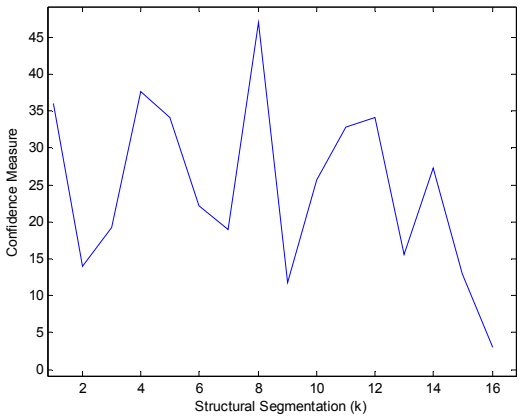

Figure 9. A graph representing the confidence measure associated with each possible segmentation. In this case, the segmentation at index 8 scored the highest with a confidence score of 47. This particular segmentation at index 8 corresponds to the structure 'AABB'.

The starting times of each part are also provided. These times correspond to the time locations of the first beat of each part.

4.6.3. The 'ABAB' case

For tunes with a time signature of 4/4, on occasion the beat tracker may detect the tempo of the audio as either double or half the actual tempo. For this reason, a heuristic is introduced especially to deal with the case when the labeling results in 'ABAB'. This particular form is very rare within this music type. It is generally only present during the performance of single reels which constitute only 0.1% of Irish tunes [15]. However, if this should possibly occur, steps are taken to correct segmentations that return this labeling as a result. Although the 'set accented tones' may actually be perceived to contain the structure 'ABAB', the likelihood is that the tempo was doubled or halved during beat tracking. If the tempo is doubled, the algorithm is expecting double the amount of parts. For this case the structure 'ABAB' actually corresponds to 'AA' when the tempo is correctly calculated. The opposite of this problem arises if the tempo is halved by the beat tracker. The automatically calculated structure 'ABAB' will correspond to the actual structure of 'AABBAABB'.

When this 'ABAB' case arises, the average length of the automatically detected structural segments of that particular tune is calculated. A weighting function that can be seen in Figure 10 is applied to this average segment length to determine whether the tempo was doubled or halved. The labeling and structural segment times for that tune are adjusted according to whichever part length is more likely according to the weighting function. The weighting function peak of 10.6 seconds was calculated by assuming a bpm of 180 for a tune of 8 bars per part. A bpm of 180 was chosen due to the very quick, up tempo nature of Irish Traditional Music.

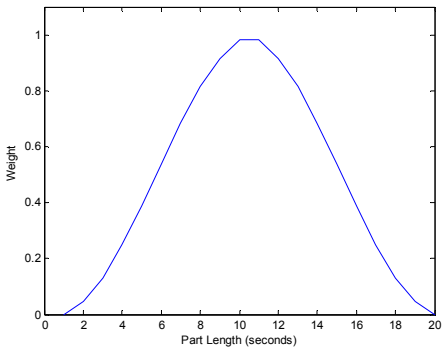

Figure 10. Weighting function used to determine actual part length for the 'ABAB' case. The weighting function peak lies at 10.6 seconds.

5. RESULTS

In order to test the performance of the segmentation algorithm, 44 tunes from various commercial recordings of solo Irish Traditional Music were used. The audio was sampled at 44100 Hz, 16-bit mono. A ground truth segmentation was manually generated by an Irish Traditional Musician in order to evaluate the algorithm. This ground truth includes the times within the audio where a structural segment change occurs and a labeling for the resulting structural segments. Automatically detected structural segment times were considered acceptable if they were within 1 second of the equivalent hand annotated structural segment times.

The results of the algorithm are outlined in Table 1. Accuracy was measured for the two separate problems of detecting structural segments and labeling those structural segments.

	Parts	*GP*	*FP*	*FN*	*pGP*	*pFP*	*pFN*	*Acc*
Times	266	210	60	45	79%	24%	18%	61%
Labels	266	192	74	60	72%	29%	24%	50%

Table 1. Results of the structural segmentation algorithm. *GP* corresponds to Good Positives, *FP* is False Positives, *FN* is False Negatives, and *Acc* corresponds to Accuracy.

Accuracy was calculated using the following formula where N is equal to the number of annotated structural segments:

$$Acc = \frac{N - FP - FN}{N} \qquad (4)$$

The value *pGP* from Table 1 is the percentage of Good Positives out of the total annotated parts. The values *pFP* and *pFN* from Table 1are the percentages of False Positives and False Negatives out of the total number of detected parts.

It is worth noting that of the 44 tunes tested, 68% of these yielded a complete and accurate segmentation and labeling. The errors that can be seen in Table 1 are all resulting from the remaining 32% of tunes. 10 of the tunes that contained errors were randomly selected and

subjected to further testing to identify the origins of the errors.

The segmentation algorithm is dependant on the accuracy of the beat tracker and the pitch detection. Consequently, if either one of these steps performs poorly for a particular audio track the ability of the segmentation algorithm to provide accurate results is adversely affected. To test the segmentation algorithm independently from the beat tracker and the pitch detector the beats from these 10 tracks were manually annotated by an Irish Traditional Musician. Subsequently, 'set accented tones' were manually extracted and inputted directly to the structural segmentation section of the algorithm, bypassing the beat tracker and the pitch detector. Patterns were then searched for amongst the ground truth 'set accented tones' vectors to determine the ability of the structural segmentation algorithm. The results obtained from testing these 10 tracks are outlined in Table 3.

	Parts	*GP*	*FP*	*FN*	*pGP*	*pFP*	*pFN*	*Acc*
Times	62	56	0	6	90%	0%	11%	90%
Labels	62	51	5	11	82%	9%	20%	74%

Table 3. Results of the structural segmentation independent from the beat tracker and pitch estimator.

The improvement on the results contained in Table 3 from the results contained in Table 2 suggests that using 'set accented tones' is an appropriate approach for detecting structural segments within musical audio.

6. CONCLUSIONS

An approach has been presented that structurally segments Irish Traditional Music in to its constituent parts. The approach used 'set accented tones' to represent each tune in its entirety. The 'set accented tones' were determined by applying a beat tracker to the audio followed by pitch detection applied to each specific beat location. Repeating patterns were then searched for amongst the resulting vector of 'set accented tones'. The structural segments were also labeled according to the similarities amongst the structural segments. Heuristics specific to this genre were used to produce more accurate results.

Results show the approach to be 61% accurate when detecting structural segment points when an automatic beat tracker and pitch detector was employed. The labeling of the structural segments was calculated to be 50% accurate as classification of segments is by nature a more complex problem than segmentation on its own. When tested independently from the beat tracker and pitch detector using ground truth 'set accented tones', the segmentation algorithm was shown to be 90% accurate at detecting the structural segment times and the labeling of those structural segments was calculated to be 74% accurate.

Future work will include adapting the algorithm to structurally segment and distinguish between multiple tunes within the same audio track. Testing and adapting the algorithm for use with polyphonic recordings also warrants future work.

7. ACKNOWLEDGEMENTS

This work was carried out as part of the IMAAS project funded by Enterprise Ireland under the Commercialisation fund.

8. REFERENCES

[1] Foote, J. *Visualizing Musical Structure and Rhythm via Self-Similarity*. in *ACM multimedia*. 2001. Orlando, Florida.

[2] Meudic, B. *Musical Pattern Extraction: from Reptition to Musical Structure*. in *Computer Music Modelling and Retrieval*. 2003. Montpellier, France.

[3] Jensen, K., *Multiple Scale Music Segmentation Using Rhythm, Timbre, and Harmony.* EURASIP Journal on Advances in Signal Processing, 2007.

[4] Ong, B.a.H., P. *Semantic Segmentation Of Music Audio Contents*. in *International Computer Music Conference*. 2005. Barcelona, Spain.

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

[6] Logan, B. *Music Summarization Using Key Phrases*. in *International Conference on Audio Speech and Signal Processing*. 2000. Istanbul, Turkey.

[7] Dannenberg, R.B.a.N.H., *Discovering Musical Structure in Audio Recordings*. Lecture Notes in Computer Science. Vol. 2445. 2002: Springer Berlin.

[8] Levy, M.a.S., M. and Casey, M. *Extraction of High-Level Musical Structure From Audio Data and Its Application to Thumbnail Generation*. in *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2006. Toulouse, France.

[9] Hung-Chen, C., L. Chih-Hsiang, and A.L.P. Chen. *Music segmentation by rhythmic features and melodic shapes*. in *IEEE International Conference on Multimedia and Expo*. 2004. Taipei, Taiwan.

[10] Bartsch, M.A. and G.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*. 2001. New York.

[11] Goto, M. *A Chorus-Section Detecting Method for Musical Audio Signals*. in *IEEE Conference on Acoustics, Speech, and Signal Processing*. 2003. Hong Kong.

[12] Jian, M.-H., C.-H. Lin, and A.L.P. Chen. *Perceptual analysis for music segmentation*. in *Storage and Retrieval Methods and Applications for Multimedia 2004*. 2003. San Jose, CA, USA: SPIE.

[13] Ó'Súilleabháin, M., *Innovation and Tradition in the Music of Tommie Potts*. 1987, Queen's University: Belfast.

[14] Gainza, M., Barry, D. and Coyle, E., *Dynamic Audio Beat Tracking*. Technical Report, TR020208. 2008.

[15] Ng, A. *Rhythm Distribution in Recordings (All Recordings)*. http://www.irishtune.info/rhythm/distribution-recordings-all.html. 2000. [cited 2007 10th Dec].