MEMORY AND MELODIC DENSITY: A MODEL FOR MELODY SEGMENTATION

Miguel Ferrand, Peter Nelson

University of Edinburgh
Scotland, UK
{mferrand,pwn}@music.ed.ac.uk

Geraint Wiggins

City University
London, UK
geraint@soi.city.ac.uk

ABSTRACT

We present a memory-based model for melodic segmentation based on the notion of melodic density. The model emphasises the role of short-term memory and time in music listening, by modelling the effects of recency in the perception of boundaries. We describe the model in detail and compare it with Cambouropoulos' Local Boundary Detection Model for a series of melody examples. First results indicate that this new model is more conservative, as it generates fewer total boundaries but preserves most boundaries that coincide with the limits of recurring patterns.

1. INTRODUCTION

It is known that listeners identify segmentation boundaries when abstracting musical contents. The ability to partition a melody in several segments provides a structural description of the piece of music. Thus, segmentation can be seen as a pre-processing stage for other tasks such as pattern discovery or music search.

Pattern finding algorithms, in particular, are known to be computationally expensive, and therefore can benefit from a reduction of the initial search space. A low-level segmentation can provide an efficiency gain by pre-processing a melodic sequence, and generating an initial set of boundaries which may be used as markers for pattern search [1]. One such method is The Local Boundary Detection Model (LBDM) [2], a segmentation model that identifies discontinuities in a melodic surface based on Gestalt principles of perception. The LBDM is an essential reference amongst segmentation algorithms, mostly due to its simplicity and generality [3, 2]. As the author emphasises, the LBDM is not a complete model of grouping in itself, as it relies on complementary models (i.e. pattern similarity) to select the most relevant boundaries. Although in that context this may not be considered a weakness of the model, excessive boundary generation may become a disadvantage if we intend to use the LBDM in isolation, and when segmentation is to be used as a reliable data reduction technique.

The LBDM has a fairly short memory as it considers at most 4 consecutive events at a time. As a consequence, there is limited interaction between neighboring boundaries and sometimes small "oscillations" can be identified as salient boundaries. This type of limitation has also been referred to by Lerdahl & Jackendoff in their Generative Theory of Tonal Music [4].

Research on auditory perception and memory has underlined the influence of time in the perception of differences and in the establishment of temporal relations in sequential processes. Studies have shown that listeners retain auditory information for some time, even after the end of stimulation [5]. This means that several past (although relatively recent) stimuli may draw the listener's attention, and may be retained as the actual most recent and prominent stimuli. Some researches have suggested that listeners perceive a musical surface by focusing on successive zones, that can be viewed as a "sliding window" along the musical piece [6]. The size of this window (determined by short-term memory restrictions) should limit the amount of musical material that can be looked back on when processing a melodic sequence. Within this time window, recency effects are likely to apply, as documented in [7, 8].

2. THE LBDM

The LBDM calculates a boundary profile for a melody, using Gestalt-based identity–change and proximity–difference rules, applied to several parameters describing a melody. The refined version of this algorithm [2] takes as input a melodic sequence converted into several independent parametric interval profiles $P_k = [x_1, x_2, ...x_n]$ where $k \in \{pitch, ioi, rest\}, x_i \geq 0$ and $i \in \{1, 2, ...n\}$. A $Change$ rule assigns boundaries to intervals with strength proportional to the degree of change between neighboring consecutive interval pairs. Then a $Proximity$ rule scales the previous boundaries proportionally to the size of the intervals.

The strength of the boundaries at each interval x_i is given by the following,

$$s_i = x_i \times (r_{i-1,i} + r_{i,i+1}) \tag{1}$$

where

$$r_{i,i+1} = \begin{cases} \frac{|x_i - x_{i+1}|}{x_i + x_{i+1}} & x_i + x_{i+1} \neq 0 \wedge x_i, x_{i+1} \geq 0 \\ 0 & x_i = x_{i+1} = 0 \end{cases}$$

For each parameter k a sequence s_k is calculated, then all sequences are normalised and combined in a weighted sum to give the overall boundary strength profile. The suggested weights for the 3 different parameters are $w_{pitch} = w_{rest} = 0.25$ and $w_{ioi} = 0.5$ (see [9] for an overview on the behavior of the LBDM with different parameter tunings). The local peaks in the resulting boundary profile indicate local boundaries in the melodic sequence. A threshold must be defined a priori, above which, a peak is identified as a boundary. For additional details on the implementation of the LBDM the reader is referred to [2].

3. MELODIC DENSITY SEGMENTATION MODEL

We now describe a new model for melodic segmentation which identifies segmentation boundaries as perceived changes in melodic

Table 1: $Order$ and $recency$ of pitch intervals for a sequence of events. Intervals are in semitones.

e_{i-3}		e_{i-2}		e_{i-1}		e_i	$event$
53		50		50		48	$pitch$
							$order(n)$
	3		0		2		1
			3		2		2
					5		3
...	2		1		0		$recency(m)$

density. We will designate this model as Melodic Density Segmentation Model (MDSM). In contrast with the LBDM, that measures the accumulated boundary strength and identifies local maxima, the MDSM calculates the accumulated melodic cohesion between pitch intervals, and then identifies local minima (i.e. points of low melodic density) as local boundaries. This new segmentation method also incorporates a short-term memory window and models the effects of recency with an attenuation function.

Before a formal description of the model is presented, some of its characteristics and underlying assumptions must be explained.

It is conjectured that pitch intervals may be formed (and perceived) between all notes occurring over an interval of time (short term memory window) and not just between consecutive notes. In Table 1 a short sequence of 4 midi notes is depicted together with the pitch distances between all pairs of events. The order of an interval determines the distance between the present and previous event considered. Thus, an interval of order k with respect to a given event e_i is denoted by (e_{i-k}, e_i). For example, from table 1 intervals (e_{i-1}, e_i) and (e_{i-2}, e_{i-1}) have order 1, intervals (e_{i-2}, e_i) and (e_{i-3}, e_{i-1}) have order 2, etc...

Recency effects apply in two different ways. The higher the order of an interval, the greater the temporal separation between the events, and therefore the weaker the perceived link between the two. On the other hand, more recently formed intervals have a stronger contribution to the melodic cohesion of the sequence than earlier formed ones. The recency of an interval with respect to an event e_i is given by the time that separates e_i and the latest event of the two that constitute the interval. These two factors are combined to determine the overall contribution of each interval at any given moment in time. In Table 1, recency is indicated in the bottom row. Increasing values of recency express less recent intervals. Let's consider here for simplicity, that all events in the previous example have equidistant on-set times and equal duration. Then intervals (e_{i-2}, e_i) and (e_{i-2}, e_{i-1}) will have equivalent contribution, since the former is an interval of order 2 (meaning that events are separated by 2 duration units) but with recency 0, and the latter has order 1 but recency 1 (meaning that the interval is separated from the reference event e_i by 1 duration unit).

The melodic cohesion of an interval is defined here to be proportional to the frequency of occurrence of that interval in the interval framework associated with the melody being analysed. Later, we will discuss in more detail how these interval frequencies are obtained.

A short–term memory window determines the span of recent events that can form intervals. The size (duration) of this window is fixed. The tempo of the piece will determine the number of recent events that can be recalled and influence the perception of a boundary.

We can now formalise the notion of melodic density (MD) as the weighted sum of the contributions of all intervals occurring over a period of time determined by the memory window. So given a sequence of N events $(e_1, ..., e_N)$ representing a melodic sequence the melodic density d_i at event i, is defined as:

$$d_i = \sum_{m=0}^{t_i - t_{i-m} < M} \sum_{n=1}^{t_i - t_{i-m-n} < M} f(r_i(m,n)) \cdot a_i(m,n) \quad (2)$$

where $f(r)$ is a function that returns the frequency of an interval, and $f(r) \in [0,1], r_i \in 0, 1, ...12$, and $r_i(m,n) = |p_{i-m} - p_{i-m-n}|$ denotes a pitch interval in semitones, where p_k denotes the MIDI pitch of event e_k, and

$$a_i(m,n) = (1 - \frac{t_i - t_{i-m-n}}{M})^2 \quad (3)$$

is the attenuation function, where t_i denotes the onset time of event e_i, and M is the duration of the memory window (in seconds). It is worth noting that a Gestalt-based principle of proximity is encapsulated in the attenuation function, as this will return values closer to 1 for recent and low-order intervals, and values closer to 0 for remote and high-order intervals.

Finally, boundaries are indicated by local minima in the melodic density profile obtained from Equation 2.

4. EXPERIMENTS AND RESULTS

To assess the behavior of the model we used both the LBDM and the MDSM on a set of melody examples. For each of the examples we also obtained a pattern boundary profile, which indicates the location of recurrent patterns within the melodic sequence (see [1] for details).

The interval frequencies given by function f were obtained from the combined frequencies of intervals that occurr in major and minor scales. This major-minor framework is described by Camboroupoulos in his General Pitch Interval Representation (GPIR) [1]. The memory window M was set to 4 seconds.

Table 2 summarises the boundary counts for each melody, including pattern boundaries and the segment boundaries generated by both the LDBM and the MDSM. A boundary is marked correct if its location coincides with a pattern boundary, with a tolerance of +/-1 event. A threshold of 70% was adopted to filter only the most prominent peaks from the boundary profiles.

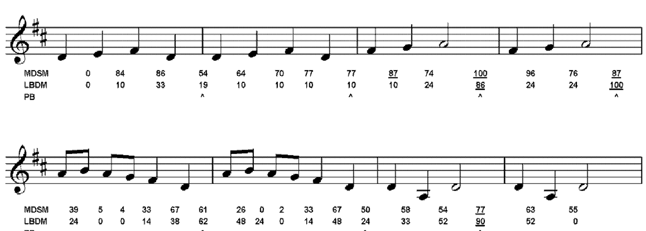

Figure 1: Normalised MDSM and LBDM boundary profiles for melody number 2 (Frere Jacques). Underlined values indicate selected peaks. Pattern Boundaries(PB) are indicated in the bottom row

Table 2: Results obtained for 7 melodies, showing the total no. of pattern boundaries (PB), and for both the LBDM and MDSM: total no. of pattern boundaries found ($_{fnd}$), no. of pattern boundaries not found ($_{notfnd}$) and no. expurious boundaries found ($_{ex}$)

Melody	PB	LBDM fnd	LBDM $notfnd$	LBDM ex	MDSM fnd	MDSM $notfnd$	MDSM ex
1. L. Row	5	5	0	0	5	0	0
2. Frere J.	7	3	4	0	5	2	0
3. Twinkle	5	5	0	2	4	1	1
4. Y.Doodle	5	5	2	3	5	0	2
5. L'H.Arme	9	8	1	0	9	0	0
6. Mozt.Gm	6	6	0	14	6	0	3
7. Beet.9th	9	9	0	0	9	0	0
Total	46	39	7	19	43	3	6

Table 3: F-measure for the LBDM and MDSM

Model	P	R	F
LBDM	0.85	0.67	0.75
MDSM	0.93	0.88	0.91

In total the LBDM generated 58 boundaries against only 49 by the MDSM. From the analysis of Table 2 it may be observed that both models find approximately the same number of pattern boundaries, but the MDSM is more conservative, generating only 6 excessive boundaries, against the 19 of the LBDM. In the melodies where excessive boundaries where found, the MDSM always register a lower count. However It must be noted that melody number 6 alone (theme of Mozart's Symphony in Gm) is responsible for the majority of the excessive boundaries generated by the LBDM. For a numerical comparison between the performance of both models the F-measure [10] was used. The F-measure is given by the weighted harmonic mean of $Precision(P)$ and $Recall(R)$.

$$F_{measure} = 2 \times \frac{P \times R}{P + R} \tag{4}$$

where

$$P = \frac{PB_{fnd}}{PB_{fnd} + PB_{notfnd}}, R = \frac{PB_{fnd}}{PB_{fnd} + PB_{excess_{fnd}}} \tag{5}$$

In table 3 we can see that although the MDSM only has a slightly higher $Precision$, it has a significantly higher $Recall$ resulting in a higher value of F.

In Figure 1 we show the boundary profiles of both models together with the score of melody no. 2 (Frere Jacques). For ease of comparison, the melodic density profile of the MDSM has been inverted [1] and normalised in the range 0-100%. From this example it seems clear that some of the boundaries generated by the LBDM were eliminated due to the 70% selection threshold, although smaller peaks can be found in the vicinity of the pattern boundaries that were missed.. An adjustment of the selection threshold to considerably lower values, will result in a significant increase of the number of peaks that are extracted, and consequently in an increase of the number of spurious boundaries. On the other hand, we would expect that an increase of the selection threshold would increase the selectivity of the model. In Figure 2 we can observe that this is not always the case. Most of the peaks of the LBDM profile have values over 80% or even 90%, thus making the elimination of the excessive boundaries difficult to achieve only by adjusting the selection threshold. The example of Figure 2 highlights also that most of the boundaries "filtered" by the MDSM are not coincident with pattern boundaries.

[1] recall that for the MDSM boundaries are obtained from the lower peaks on the profiles

5. DISCUSSION

The boundary selectivity reported on the MDSM, results partially from the propagation of the intervals over a time window creating a "smoothing" effect. However this effect can be also a drawback of this approach. In some cases, boundaries can be shifted forward or prolonged due to a slower decay of the melodic density function. This is visible in Figure 1 where the boundary peak after the third measure is followed by a significantly slow decay of the MDSM values (specially when compared with the sharp drop on the LDBM profile), until it meets the following peak. This may have an impact on the accuracy of the boundary locations, in particular when matched without tolerance, against pattern boundaries.

Although tempo was kept constant in this study, the MDSM is robust to small changes in tempo. This is mainly due to the discrete nature of the events, combined with a memory window of fixed size. For example, with a tempo of crotchet=60, a memory window of 5 seconds would include 5 crotchets (or the equivalent in duration), and an increase of the tempo to crotchet=72 would be necessary to include an additional crotchet in the calculations. Few studies have addressed the effects of changes in tempo in music perception [11]. Although the present model was designed to account for changes in tempo, a systematic evaluation of these effects has not yet been included. For such analysis we may require that listeners be tested on the effects of changes in tempo to provide data to be compared with the model.

The choice of the attenuation function (a decaying polynomial), is the result of preliminary experiments with the algorithm, where several decaying functions were examined. However, it must be said, the differences were not conclusive. It seems intuitive that, in general, less recent notes have a smaller contribution to the melodic cohesion of a sequence, than more recent ones. However, to the best of our knowledge, there is no theoretical or experimental evidence to support the choice of a specific memory decaying function.

As mentioned previously, interval frequencies were obtained from the combined statistics of interval counts from major and minor scales. Since one of the motivations of this work is to devise a model that can segment melodies without any domain specific knowledge, we propose that these frequencies may be acquired from a music corpus that is representative of the melodies being analysed. This idea is supported by several studies, some of which were carried out outside the western musical culture, that report, for example, the prevalence of small melodic intervals in melodic lines [7, 12]. If indeed the melodic preferences of a particular musical culture are reflected in the musical material, it seems rea-

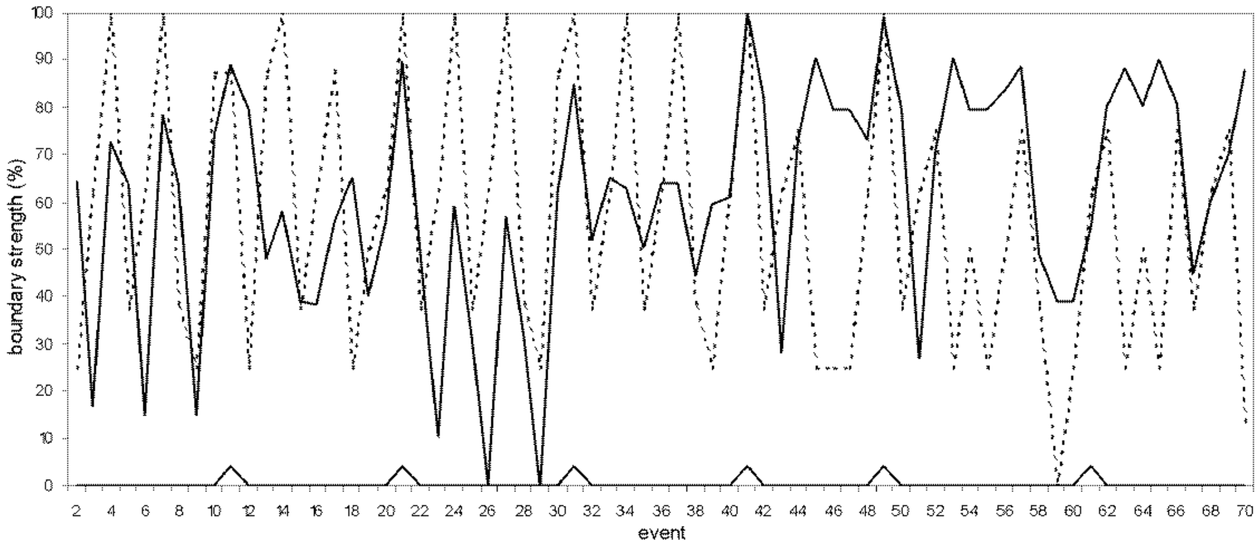

Figure 2: Boundary profiles obtained with LBDM (dotted line) and MDSM (solid line) for melody no. 6 (theme of Mozart's Symphony in Gm). Pattern boundaries are indicated by arrows at the bottom of the chart.

sonable to reverse this process, by using implicit intervalic information to interpret the musical material.

6. CONCLUSIONS

We presented the MDSM, a memory-based melodic segmentation algorithm based on the concept of melodic density. We compared this algorithm with the LBDM, for a set of melody examples. It was shown that in general the MDSM has higher selectivity than the LBDM, generating fewer total boundaries but preserving most boundaries indicated as pattern boundaries. This suggests that the MDSM may be used successfully as a pre-processing method for pattern finding algorithms, providing additional reduction of the search space without the cost of eliminating many candidate pattern boundaries.

The contribution of this new approach lies in the way it incorporates pitch and time, and in particular in the use of tempo as a parameter together with a short-term memory window, thus seeking a more cognitively realistic approach to melodic segmentation.

7. ACKNOWLEDGMENTS

This research was funded by EPSRC research project GR/N08049/01

8. REFERENCES

[1] Emilios Cambouropoulos, *Towards a General Computational Theory of Musical Structure*, Ph.D. thesis, University of Edinburgh, 1998.

[2] Emilios Cambouropoulos, "The local boundary detection model (lbdm) and its application in the study of expressive timing," in *Proceedings of the International Computer Music Conference (ICMC'2001)*, Havana, Cuba, September 2001.

[3] Emilios Cambouropoulos, "Melodic cue abstraction, similarity, and category formation: A formal model," *Music Perception*, vol. 18, no. 3, pp. 347–370, 2001.

[4] Fred Lerdahl and Ray Jackendoff, *A Generative Theory of Tonal Music*, M.I.T. Press, Cambridge (Mass.), 1983.

[5] M. Eysenck and M.T. Keane, *Cognitive Psychology: a Student's Handbook*, Psychology Press, 3rd edition, 1995.

[6] Emmanuel Bigand, "Contributions of music to research on human auditory cognition," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, Stephen McAdams and Emmanuel Bigand, Eds., chapter 8, pp. 231–277. Oxford University Press, 1993.

[7] Carol L. Krumhansl, *Cognitive Foundations of Musical Pitch*, Number 17 in Oxford Psychology Series. Oxford University Press, Department of Psychology, Cornel University, 1990.

[8] Ludger Hofmann-Engl and Richard Parncutt, "Computational modeling of melodic similarity judgments: two experiments on isochronous melodic fragments," http://freespace.virgin.net/ludger.hofmann-engl/similarity.html, 2000.

[9] Belinda Thom, Christian Spevak, and Karin Höthker, "Melodic segmentation: Evaluating the performance of algorithms and musical experts," in *Proceedings of the 2002 International Computer Music Conference (ICMC'02*, Göteborg, Sweden, 2002.

[10] C. J. Van Rijsbergen, *Information Retrieval, 2nd edition*, Butterworths, London, Boston, 1979.

[11] Stephen Handel, "The effect of tempo and tone duration on rhythmic discrimination," *Perception & Psychophysics*, vol. 54, no. 3, pp. 370–382, 1993.

[12] Paul von Hippel, "Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals," *Music Perception*, vol. 17, no. 3, pp. 315–327, 2000.