OPTIMISING PARAMETER WEIGHTS IN MODELS FOR MELODIC SEGMENTATION

Tillman Weyde
Research Department of Music and Media Technology, University of Osnabrück, Germany

ABSTRACT

Segmentation of melodies is an important issue in music psychology as well as computer applications. It has been approached by music theory and music psychology but a coherent theory or model has not yet been established. Some influential factors for segmentation are known, but it is generally not clear how they can be integrated into a computer model. This paper is concerned with the influence and interaction of some of these factors. In a first experiment with artificial melodies pairs of factors are isolated to assess the type of model needed. In a second experiment with random melodies a larger set of features is used. These used as input to two types of adaptive models to compare their performance on this task, showing that a linear model is not powerful enough.

1. INTRODUCTION

Segmentation of music by perceptual grouping is an important part of understanding music because it defines the meaningful units of musical structure. Musical computer applications need to identify these units for musical tasks like music retrieval, yet only few theories and model have been developed.

The phenomenon of subjective rhythmization has been discovered early by Bolton (1894) and Wundt (1903) and the *Gestalt* psychologists have emphasized the grouping of elements by proximity and similarity (von Ehrenfels 1890). The temporal constraints of grouping notes to motifs have be studied e.g. by Fraisse (1982) and Pöppel (1989).

This work is part of the project MUSITECH funded by the Deutsche Forschungsgemeinschaft DFG.

The first to model segmentation by computer were Tenney and Polansky (1980), who implemented a model of segmentation based on finding local maxima of an inter-note distance metric. The *Generative* Theory of Tonal Music by Lerdahl and Jackendoff (1983) contains a set of Grouping Preference and Well-Formedness Rules which have been empirically supported by Deliége (1987). Newer models include the *Local Boundary Detection Model* (LBDM) by Cambouropoulos (1996) and the *Integrated Segmentation and Similarity Model* by the author (Weyde 2002). Some factors have been shown to influence listener's segmentation preferences (see e.g. Bregman, 1990). Yet the formal and computer based models revealed that knowing these factors is not sufficient to develop a model, it is yet unclear how the combination and interdependence of different factors and perceptual constraints are to be modelled.

2. INFLUENTIAL FACTORS FOR SEGMENTATION

Segmentation is influenced by many factors: rhythmic, melodic, metric, motivic, harmonic etc. Here we will only discuss factors that are directly related to the segments, not those based on their motivic relations by similarity and no metrical aspects.

Motif Length and Duration The length (number of notes) and the duration (temporal extension) of motifs are constrained by the properties of auditory perception. According to Fraisse (1982) and Pöppel (1989) the maximum duration of perceptual groups lies in the range of 3 seconds, although in a musical context they are usually shorter. The lower bound of a musical motif duration is not as clearly known, but there is evidence that it is in the range of 250-500 ms (Seifert et al., 1995).

The maximum length of a motif is constrained by the capacity of short term memory, which is about 7 elements according to Miller (1956). But already with more than 4 elements the recall of element order decreases, and in a musical context motifs of more that 4 notes are rare (Handel, 1989). Groups with only one note are also rare and should be avoided according to GTTM.

Regularity It has been found that listeners prefer segmentations which are regular in respect to length and duration (Handel and Todd, 1981). Groupings which correspond to metrical structure are also preferred, but this will not discussed here.

Proximity Proximity is important factors for segmentation as temporal distances introduce group boundaries(see Deutsch, 1986). As Tenney and Polansky (1980) noted, similarity can take the same mathematical form as proximity if we can reduce similarity to one dimension by a metric. This corresponds to using spatial metaphors for temporal and tonal dimensional (a *long high* note). So it seems natural to think of pitch and time as a two-dimensional space. One may include loudness, but it seems that loudness has a different effect: large distances in pitch or onset time tend to mark a group boundary while with loudness this is only true if it is rising not if it is falling (Bregman, 1990).

Gestalt and Similarity Another parameter is the direction of movement, which gives a sense of continuity when it is constant and indicates a group boundary when it changes (Bregman, 1990). There are of course more Gestalt parameters related to motifs and their relations but they will not be covered here.

ISBN 3-931852-67-9
ISSN 1617-6847

Figure 1: Piano roll visualization of Pitch/IOI example with pitch distance 5 semitones and IOI 150 (horizontal lines mark every semitone, vertical lines every second).

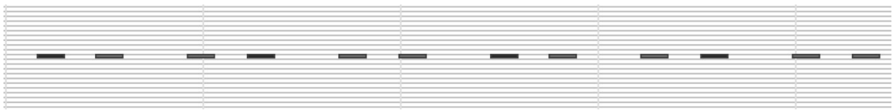

Figure 3: Piano roll visualization of loudness/IOI example with louder notes displayed darker.

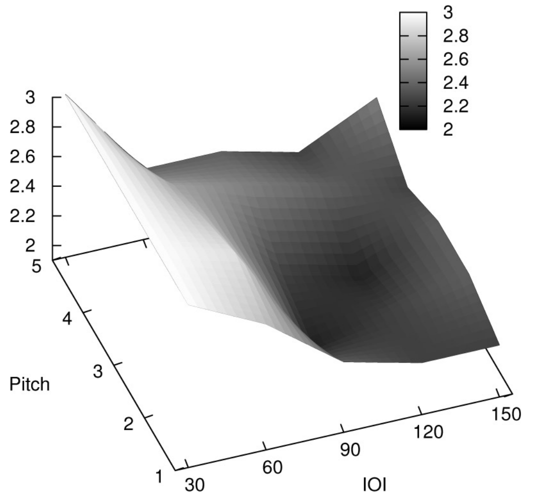

Figure 2: Interpolated average segment length in the pitch/IOI experiment.

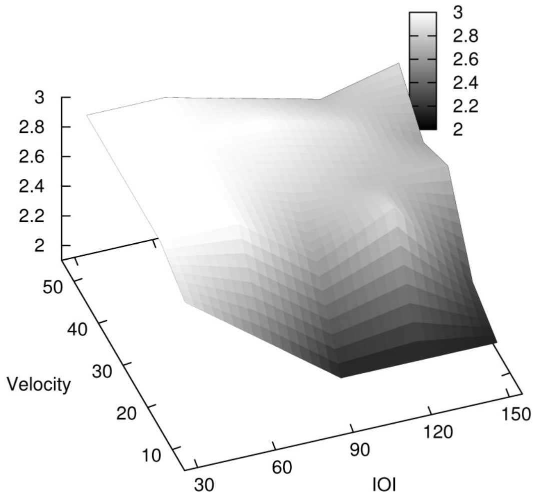

Figure 4: Interpolated average segment length in loudness/IOI experiment.

2.1. Experiment 1

The main principle of most segmentation theories is to put segment boundaries at local maxima of pitch and time intervals. The combination of different parameters like time and pitch is modelled by scaling them before combining them in a metric. Cambouropoulos (1996) additionally uses the change of intervals and his LBDM but also relies on scaling the different parameters. To see how the different parameters are related an experiment with three types of note sequences was conducted. These sequences have been constructed to vary two parameters independently of which one leads to segmentation into groups of two and the other into groups of three notes. The two parameters were varied in 5 steps independently in a range which leads to change of perceived segmentation. The generated 25 sequences were then played to eight subjects (music students) who were asked if they preferred a grouping in motifs of two or three notes.

Pitch/IOI In the pitch/IOI experiment note sequences were created which had a change in pitch every three notes alternating up and down. This change was 1 to 5 semitones. An extended inter onset interval (IOI) followed after every second note where the extensions were 30, 50, 90, 120, and 150 ms. An example is shown in figure 1.

The average choice of the subjects is shown in figure 2. The result shows that the IOI is more influential than the pitch interval, at least in the range covered here. This supports the view of Deutsch (1986) who supposes a dominance of temporal proximity over other factors. The dependence on IOI does not seem to be linear.

Loudness/IOI The second part of the experiment used the same setting as before but loudness instead of pitch as MIDI velocity differences between 10 and 50 in steps of ten 2. The results are shown in figure 3. Here IOIs and loudness contribute to change in segmentation group length.

Direction/Interval In this experiment we have a series of notes with pitch changed by a constant interval (internal interval) which changes the direction every three notes and contains a pitch interval of enlarged size every second note (external interval). So the larger the interval, the steeper direction of movement and the larger the external interval the larger the disruption of the movement. An example is shown in figure 4. The result is shown in figure 5. Here the surface is rugged which indicates a nonlinear relation.

ISBN 3-931852-67-9
ISSN 1617-6847

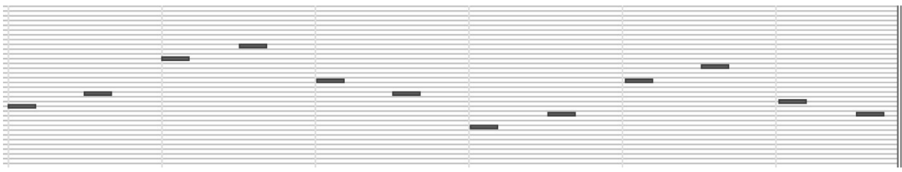

Figure 5: Piano roll visualization of direction/interval example.

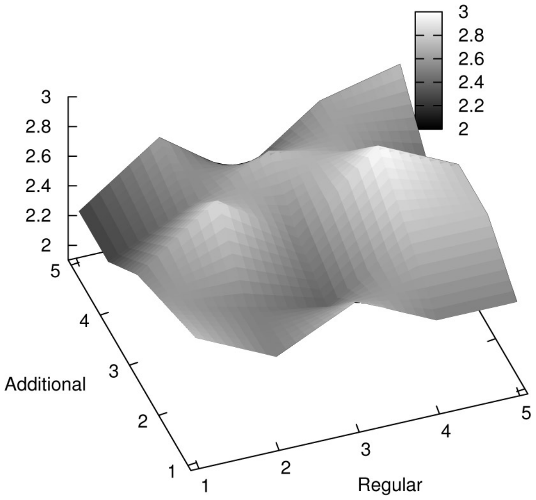

Figure 6: Average segment length in direction/interval experiment.

Results The results of this small experiment indicate that a linear model may be inadequate for modelling the combination of segmentation parameters. This may be due to high variability and small data sets but from these data it seems unlikely that a linear model could provide a good fitting to real data. To gain more certainty on this point a second experiments follows.

3. MODELLING SEGMENTATION OF MELODIES

When realising melodic segmentation on a computer the influence and interaction of parameters, modelled by parameter weights, have to be adjusted. Adaptive systems are useful here because a they provide way to fine-tune the model parameters by examples.

The system used for performing segmentation is the *Integrated Segmentation and Similarity Model*, only the segmentation part of which is used here. Its functions can only be outlined here, more details can be found in Weyde (2001, 2002). It creates all possible segmentations and chooses the one that receives the best rating. The calculation of the rating from the input features can be any adaptive mapping system that can be trained by backpropagation. The system can be optimized by examples using *Iterative Training*. This means that for every example it is tested if the system delivers the same output as given in the example. If not, a relative sample is created. The training by relative samples uses backpropagation to change the ratings of the example and the system output so that the example will be rated better that the system output. This procedure is iterated until it either converges or meets another termination criterion (e.g. number of iterations). The following input features were used in addition those mentioned before.

- per motif:
  - ratio of largest inner IOI to outer IOIs
  - ratio of largest inner rests to outer rests
  - ratio of largest inner pitch interval to outer intervals
  - interval ratio by distance in circle of fifth
  - change in pitch direction
- per sequence:
  - regularity of motif length (number of notes per motif)
  - regularity of motif duration (from motif to next motif)

Numbered lists are useful when an order or ranking is implied in the list items.

3.1. Experiment 2

In the second experiment randomly generated sequences were used. Sequences which were subjectively judged hard to interpret musically were removed. 20 of the remaining sequences consisting of six to seven notes were used. This set was randomly divided into two sets of ten. They were presented to six subjects (music students) who were asked to describe an adequate segmentation at the lowest grouping level for each sequence of both sets. The collected segmentations were used as samples for training the ISSM. Two different adaptive systems were used: a linear model and a neuro-fuzzy system as described in Weyde and Dalinghaus (2001). They were trained with one sample set and tested on the other.

3.2. Results

Segmentations by the subjects differed largely and the system could not be trained successfully using the data directly. Closer examination revealed that many segmentations differed only in the level of subdivisions, e.g. a four-note motif was split into two motifs of two notes. So such segmentations were defined as *compatible*:

ISBN 3-931852-67-9
ISSN 1617-6847

Compatible Segmentations
Two segmentations Sg_1, Sg_1 of the form $Sg_1, = M_1, \ldots, M_m = \{\{n_1, \ldots, n_{m_1}\}, \ldots, \{n_{s-m_m}, \ldots, n_{m_m}\}\}$ of a sequence $S = \{n_1, \ldots, n_S\}$ are called compatible if for all $M_1^i \in Sg_1$ one of the following holds:

1. $\exists M_2^j \in Sg_2$ where $M_1^i = M_2^j$
2. $\exists M_2^j, M_2^k \in Sg_2, |k-j| = 1$ where $M_1^i = M_2^j \cup M_2^k$
3. $\exists M_2^j \in Sg_2, M_1^l \in Sg_1, |l-j| = 1$ where $M_1^i \cup M_1^l = M_2^j$

Compatible segmentations were regarded as correct during training, i.e. no relative sample was generated when the system produced a different but compatible segmentation. Training with tolerance for compatible segmentations yielded the results shown int table 1. The training set contained 12 and the test set 5 inconsistent samples, i.e. samples that contained segmentations incompatible with another segmentation of the same sequence. These define lower limits on the error numbers and were subtracted from the actual numbers to make them comparable.

	Linear System	Neuro-Fuzzy System
Error on training set (more than 12)	2	1
Error on test set (more than 5)	4	0

Table 1: Training and test results for segmentation of random melodies (60 samples per set).

4. CONCLUSIONS

Parameter weights in melodic segmentation can be optimised by an adaptive system. The neuro-fuzzy system used here clearly outperforms the linear model. To achieve training results at all, it was necessary to introduce some tolerance by the definition of compatible segmentations. The two experiments showed that there is a complex interaction of the musical factors which can be modelled by using an adaptive system that optimises its parameters by examples. The weights of the trained system are partly not in accordance with the literature, which may be due to the particular set of examples. To get reliable evidence on this point it is necessary make experiments with larger data sets and to use a Bayesian model allowing to better interpret the results of training.

5. REFERENCES

1. Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, 6:145-238.

2. Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press, Cambridge, Mass.

3. Cambouropoulos, E. (1996). A formal theory for the discovery of local boundaries in a melodic surface. In *Proceedings of the IIIme Journées d'Informatique Musicale*.

4. Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception*, 4(4):325-60.

5. Deutsch, D. (1986). Auditory pattern recognition. In Boff, K. R., Kaufman, L., and Thomas, J. P., editors, *Handbook of Perception and Human Performance:Cognitive Processes and Performance*, volume 2, chapter 32, pages 32-1-49. John Wiley and Sons, New York.

6. Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D., editor, *The Psychology of Music*, chapter 6, pages 149-180. Academic Press, New York.

7. Handel, S. (1989). Listening: An Introduction to the Perception of Auditory Events. MIT Press, Cambridge, Mass.

8. Handel, S. and Todd, P. (1981). Segmentation of sequential patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 7(1):41-55.

9. Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. The MIT Press, Cambridge, Mass.

10. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81-97.

11. Pöppel, E. (1989). The measurement of music and the cerebral clock: A new theory. *Leonardo*, 22(1):83-89.

12. Seifert, U., Olk, F., and Schneider, A. (1995). On rhythm perception: Theoretical issues, empirical findings. *Journal of New Music Research*, 24(2):164-95.

13. Tenney, J. and Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, 24(2):205-41.

14. von Ehrenfels, C. (1890). über Gestaltqualitäten. *Vierteljahresschrift für wissenschaftliche Philosophie*, 14:249-292.

15. Weyde, T. (2001). Grouping, smilarity and the recognition of rhythmic structure. In *Proceedings of the International Computer Music Conference 2001*, Havanna, Cuba.

16. Weyde, T. (2002). Integrating segmentation and similarity in melodic analysis.

17. In *Proceedings of the International Conference on Music Perception and Cognition 2002*, Sydney. University of New South Wales.

18. Weyde, T. and Dalinghaus, K. (2001). Recognition of musical rhythm patterns based on a neuro-fuzzy-system. In Dagli, C. H., editor, *Proceedings of the ANNIE 2001 conference*, St. Louis, USA.

19. Wundt, W. (1903). *Grundzüge der physiologischen Psychologie*, volume 3. Wilhelm Engelmann, Leipzig, 5th edition.

ISBN 3-931852-67-9
ISSN 1617-6847