Integrating Convexity and Compactness into the ISSM: Melodic Analysis of Music

Aline K. Honingh and Tillman Weyde

Department of Computing

City University London

Technical Report Series

TR/2008/DOC/03

ISSN 1364 4009

Integrating Convexity and Compactness into the ISSM: Melodic Analysis of Music

Aline Honingh[*] and Tillman Weyde

Music Informatics Research Group,
Department of Computing
City University, London

Abstract

This paper reports on a short project that aimed to integrate two mathematical measures, convexity and compactness, into the ISSM, a model for music analysis. The measures convexity and compactness have been successfully used before in music informatics research. It turned out that the combination of these two tools has not been as successful as initially hoped for, and therefore the project has been aborted. We report here on the methods and give a rationale for not pursuing this project any further.

1 Introduction

For musicians it is in general clear where to breath or to pause after a musical phrase, when playing a piece of music. Also, trained music theorists have no problem analyzing a piece of music melodically, the repetitions and similar phrases are clear to them when they hear the piece. However, if you would ask these people why they pause at a certain moment or how they come to a certain melodic analysis, it is difficult to extract a sufficient rule system from their answers. The processes of melodic analysis circumscribe the discovery of melodic structure of a piece of music.

These tasks are of the kind that typically people can do well, but that is very difficult to model with a computer. Modeling these tasks is of twofold interest. On the one hand the automatic analysis of these musical structures can by automatic creation of semantic annotations serve as a useful tool for musicology, music information retrieval, and music education. On the other hand, there is a the possibility to gain insight in the cognitive processes involved in music by developing and analyzing a successful computational model.

The aim of this project is to develop a computational model for melodic structuring. This problem several interdependent sub-problems, such as melodic segmentation, and melodic similarity.

1.1 Current state of research

Nowadays, the problem of automatic melodic analysis has been looked at from many different sides, and various methods ranging from topological models (Buteau 2005) to statistical approaches (Bod 2002) have been used. Although several models can successfully measure similarity of two melodies having the same length, or determine the key of a short standard piece of music, the extension of these concepts and integration into a full melodic analysis model continues to be a difficult problem.

[*]Corresponding author address: Aline.Honingh.1@soi.city.ac.uk

This is the first time that the geometrical notions of convexity and compactness are applied to this problem. It has been shown that the principle of convexity is beneficial for harmonic reduction and segmentation on the basis of modulation (Honingh and Bod 2005; Honingh 2006b) and that the principle of compactness is useful in analyzing the intonation of chords and a pitch spelling model (Honingh 2006a, 2006c).

1.2 Research methods

Computational routines to measure convexity and compactness are written such that they can be applied to real music, having encoded music as input, and annotated musical structures as output. Expecting that the principles of convexity and compactness will not be sufficient to model complex musical structures, we intend to integrate these principles within the framework of ISSM (Weyde 2002; Weyde and Dalinghaus 2003). That is, besides the input features for melodic segmentation, such as the number of notes, duration and pitch intervals at motif boundaries, the features compactness and convexity could be added to strive for melodic segments that represent convex and highly compact structures.

2 Description of ISSM

The *Integrated Segmentation and Similarity Model (ISSM)* is an approach to analyze melodies that combines multiple musical aspects using machine learning (Weyde and Dalinghaus 2003; Weyde 2002). In the analysis of music there are typically two important aspects, the segmentation of the melody into structural units (called phrases or motives) and the similarity of these segments, which in combination can yield what musicologists call a motivic analysis.

The ISSM has two modes of operation: the comparison mode and the analysis mode. In comparison mode two different melodies are segmented and segments of one melody, the input, are assigned to similar segments in another melody, the model. This mode was developed for applications in music education, where the input by a student should be compared to a model, defined from a score and/or a expert performance.

In analysis mode, one melody is used as both the input and the model. This is useful for general motivic analysis. The procedure, which will be explained in more detail below, is basically the same as in comparison. However it is necessary to disallow the assignment of a segment to itself, as this would always yield highest similarity ratings and will not be musically meaningful. Also it is useful to only allow assignments of segments in one temporal direction. Usually only the later segments are allowed to be assigned to earlier ones, but not vice versa. This simplifies the analysis process without loosing relevant information, as bi-directional assignments would not give additional interpretations, but be perceptually implausible.

Within the assigned segments, individual notes of the input are assigned to notes of the model, allowing detailed analysis, e.g. for feedback to a study when indicating differences between the expected and actual performance.

For the application phase, the ISSM uses mainly a brute-force approach. It generates all possible segmentations of input and model (which have to be identical in analysis mode, when input and model are identical), and all possible assignments of the segments. A combination of segmentation(s) with assignment is called an interpretation. Every interpretation is rated for its quality using a neuro-fuzzy-system, and the interpretation with the highest quality is selected for the output.

For the note assignments, also the best assignment is chosen, but this is only done on a local basis per segment assignment. Not all combinations of note assignments for different segment assignments are tested. Keeping the search local makes the computation tractable for relevant melody lengths of 6 to 10 notes, and seems perceptually plausible (although there is to our knowledge no clear empirical evidence available on this question).

The neuro-fuzzy rating defines a set of fuzzy logical rules that operates on feature values derived from the interpretation data. Features are based on time intervals, pitch intervals, loudness differences etc. and can be associated with a segment (e.g. the ratio of the length of the rest after the segment to the rests within the segment), the whole segmentation (e.g. the degree of variance of the segment durations), or the assignments (e.g. the pitch differences between the assigned notes). The rules based on these features combine feature values on the segment and interpretation level, using a list processing approach, and specifically designed fuzzy-logical operators, that allow a variable number of arguments.

The definition of the features and the rules is based on expert knowledge, but the rules can also be defined in a canonical way (e.g. all inputs to all outputs), allowing a 'neutral' starting state for the machine learning). In addition, expert knowledge can be used indirectly, by providing the ISSM with example analyses. The ISSM uses a specifically designed learning scheme to adapt itself toward producing the same output as in the given examples. The ISSM has been tested and can actually learn to adapt itself to given examples of music analysis. Evaluating the change of rule weights through the learning process, can give information about the contribution of different rules in the segmentation and similarity.

3 Description of convexity and compactness

Convexity and compactness are notions from geometry, and have been applied to music several times for different purposes (Honingh and Bod 2005; Honingh 2006b,2007, 2008b, 2008a). To explain these two measures, we first need to introduce a tonal space, in which they can be defined.

Musical intervals can be expressed in terms of frequency ratios. Since any positive integer a can be written as a unique product $a = p_1^{e_1} \cdot p_2^{e_2} \dots p_n^{e_n}$ of positive integer powers e_i of primes $p_1 < p_2 < \dots < p_n$, any frequency ratio a/b $(a, b \in \mathbb{Z})$ can be expressed as

$$2^p 3^q 5^r \dots \tag{1}$$

with $p, q, r \in \mathbb{Z}$. For example $2^{-1}3^1 (= \frac{3}{2})$ represents a perfect fifth and $2^{-2}5^1 (= \frac{5}{4})$ a major third. Tuning according to whole number ratios is referred to as just intonation. If the highest prime that is taken into account in describing a set of intervals is n, then this is called n-limit just intonation. Focusing on 5-limit just intonation, all intervals can be described by $\{2^p 3^q 5^r | p, q, r \in \mathbb{Z}\}$ or, equivalently by

$$\{2^p (\frac{3}{2})^q (\frac{5}{4})^r | p, q, r \in \mathbb{Z}\}, \tag{2}$$

meaning that every interval can be built from a number of major thirds, perfect fifths, and octaves. The intervals from 5-limit just intonation within one octave can be represented in the two-dimensional lattice $\mathbb{Z}^2$, by making a projection

$$\varphi : 2^p (\frac{3}{2})^q (\frac{5}{4})^r \rightarrow (q, r). \tag{3}$$

The resulting space is shown in figure 1a and is referred to as Euler-lattice of frequency ratios. The frequency ratios in the Euler lattice can be labeled with note names if a reference note is

	25/18	25/24	25/16	75/64	225/128		
40/27	10/9	5/3	5/4	15/8	45/32	135/128	
32/27	16/9	4/3	1	3/2	9/8	27/16	81/64
256/135	64/45	16/15	8/5	6/5	9/5	27/20	81/80
	256/225	128/75	32/25	48/25	36/25	27/25	

(a)

	F#	C#	G#	D#	A#		
G	D	A	E	B	F#	C#	
Eb	Bb	F	C	G	D	A	E
Cb	Gb	Db	Ab	Eb	Bb	F	C
	Ebb	Bbb	Fb	Cb	Gb	Db	

(b)

	6	1	8	3	10		
7	2	9	4	11	6	1	
3	10	5	0	7	2	9	4
11	6	1	8	3	10	5	0
	2	9	4	11	6	1	

(c)

Figure 1: Euler lattice of (a) frequency ratios, (b) note names, and (c) pitch classes.

chosen and identified with the prime interval 1. In turn, all note names can be identified with pitch class numbers: the numbers 0 to 11 representing all 12 semitones within one octave. Here we have chosen to project frequency ratio 1 onto note name C (see fig. 1b) and in turn onto pitch class 0 (see fig. 1c). The full projections are obtained by projecting the generating elements 3/2 and 5/4 onto the G and E for the note names, and onto the elements 7 and 4 for the pitch classes. With these unit elements the rest of the elements can be obtained via vector addition. In this way, two homomorphic projections from the frequency ratios to the notes names (fig. 1b) and pitch classes (fig. 1c) arise. The projections from frequency ratios to note names and pitch classes are homomorphic but not isomorphic (not invertible), and boundary conditions exist on the 2-D lattices that visualize the note names and pitch classes (Honingh 2006b). Identifications of points on the lattice can be made such that the note names can be visualized on the surface of a cylinder and the pitch classes on the surface of a torus (doughnut shape) (Honingh 2006b). These three tonal spaces can be used to visualize tonal structures such as scales and chords.

Convexity on a two dimensional lattice has been defined as follows: A set is convex if all elements (lattice points) that lie in its convex hull are included in the set. In other words: A set is convex if, drawing lines between all points in the set, all elements that lie within the spanned area are elements of the set (Honingh and Bod 2005).

Compactness is intuitively understood as the degree to which elements of a set are close together in a set[1]. In previous applications of compactness to music, the compactness is calculated by summing the distances between all pairs of points (in the Euler lattice); the lower the resulting value, the more compact is the set (Honingh 2008a; Honingh 2008b).

These definitions of convexity and compactness on curved surfaces like a torus or cylinder give some complications since there is always more than one way to go from one point to another (via either side of the cylinder or torus). The cylinder or torus has to be 'cut open' (mentally) at a chosen position in order for the definitions given above to apply. Convexity is a boolean valued function (i.e. a set is convex or not), and compactness a continuous valued function (i.e. several degrees of compactness exist). The chosen convention for these measures on curved surfaces is as follows: 1. a set is convex if there exists a way to cut the curved surface open such that the set is convex according to the above definition of convexity on a two dimensional lattice, and 2. the compactness of a set on a curved surface equals the most compact set (according to the definition of compactness on a two dimensional lattice) considering the various values of compactness belonging to the various ways of cutting the surface open (Honingh 2006b).

Convexity has been presented as a measure of well-formedness of scales and chords and as a tool for automatic modulation finding, and compactness has been presented as a useful tool which forms the basis of the processes of pitch spelling and selecting the preferred intonation of chords.

4 Integrating convexity and compactness into the ISSM

When the ISSM is used for melodic analysis it operates in the analysis mode (as described in section 2). It is designed to find the most plausible segmentation and the similarity structure by generating all relevant interpretations and choosing the best interpretation. As explained above (section 2) knowledge can be integrated in the system by defining features and rules on those features. By including the measures of convexity and compactness into this list of features, the results of the ISSM can possibly be improved. It is hypothesized that compactness could help to indicate similarity: if two segments have equal (or similar) compactness they are similar. Furthermore, it is hypothesized that convexity could help to indicate segment/grouping boundaries. It has been shown that convexity was successful in identifying modulations (i.e. clusters of non-convex sets turned out to identify modulations (Honingh 2007)), and therefore it is argued that this could contribute to the identification of grouping boundaries since at the point of a modulation usually a grouping boundary exist[2].

[1]Compactness in this respect should not be confused with the topological notion of compactness

[2]It does not matter that this is not always true the other way around (i.e. at the position of a grouping boundary there is not always a modulation present) since the feature 'convexity' is part of a whole list of features

To integrate the features compactness and convexity into the ISSM, Java classes and methods have been written. The compactness of two musical segments is calculated (using the definition of compactness given in section 3) and compared using a fuzzy normalization function such that the output is a number between 0 and 1. The musical segments are sequences of note names so the calculation of the compactness of these segments applies to the tonal space presented in figure 1b. The convexity of several segments of an input stream of music is calculated a number of times with different values for 'segment size' such that the locations where a non-convex set is located can be compared and weighted. More weight is added for non-convex sets consisting of more notes. This algorithm has been used originally to locate modulations (see Honingh 2007). We argue that at the location of a modulation, there will most likely be a segment boundary (but not necessarily the other way around). The method has shown to work better for music in major mode than for music in minor mode. This algorithm finally produces a sequence of likelihood values that indicates the likelihood for a segmentation being at a certain point in the music, for all points (notes) in the input. Again, the musical segments are sequences of note names so the calculation of the convexity of these segments applies to the tonal space presented in figure 1b.

5 Results and conclusions

The musical test data for our extended model was not so easy to find. MIDI data cannot be used, since the convexity and compactness methods use the note names (as opposed to pitch classes which are related to MIDI) and distinguish between enharmonically equivalent notes[3]. Unfortunately, not so much music is available in formats that encode note names. The second difficulty is that the data to be used has to contain annotated (expert) segmentations such that we are able to compare the output of the model with a 'ground truth'. The process of finding a test corpus for our model was further complicated since the ISSM works only for small segments (i.e. segments on the group level, not the phrase level (Lerdahl and Jackendoff 1983)) and only for monophonic music (as opposed to polyphonic music). The Essen Folk song collection (Schaffrath 1995) seemed to be the single test corpus that perfectly suited our needs. The notes are encoded in solfege numbers so the note names can be easily extracted. The songs are annotated and have a 'hard return' at a phrase ending and the phrases are relatively short.

However, more unforeseen complications existed. The fact that the ISSM can only model short monophonic musical groups clashes with the fact that the convexity algorithm is optimized for long polyphonic phrases (since it is easier to locate a modulation using many notes than using only a few notes). In an attempt to optimize the convexity algorithm such that it would give an indication of the phrase boundaries of the Essen songs, we have studied many of the Essen folksongs to see what the phrase boundaries typically look like, and to discover whether we could find anything that would help us to indicate the boundaries on the Euler lattice. It turns out that many phrases end on the same note as the next phrase start with: this is in contradiction to the prediction of the convexity algorithm, since an extra instance of the same note does not change the convexity (or compactness) on its own because the cardinality of notes is not taken into account. The procedure of taking into account cardinality by adding weights if a note appears more than once has been investigated but seems to make no difference. Furthermore, phrases end often with a rest (a zero in Essen code) but rests are not covered by the convexity and compactness methods.

Preliminary tests have found that indeed the convexity algorithm cannot locate segment boundaries in the Esses folk songs: every possible set turns out to be convex (while non-convex sets would indicate a segment boundary) since generally all Essen songs are in one and the same key. Also the measure of compactness has not been found to be a good indicator of either segment boundaries or similarity.

and is therefore not meant to cover the whole concept of grouping boundaries on its own.

[3]It is important that the convexity method uses note names instead of pitch classes. Since the Euler lattice of pitch classes exists of only 12 elements in total, almost any set in this space will turn out to be convex.

The music problems that have been addressed with algorithms using convexity and compactness so far, all had a component of harmony. The Euler-lattice, also known as 'harmonic network' can adequately describe harmonic structures but has almost never been used in melody related problems. One can see that consecutive tones in the major and minor scale (important for melody) are maximally far apart in the scale on the Euler lattice, as opposed to the thirds (important for harmony) which are maximally close to each other in the scale on the Euler lattice. The problem that we address here is, mainly because we consider monophonic music, primarily melody related (as opposed to harmony related) and therefore the Euler lattice may not be the ideal tonal space to use in this case. We may conclude that, although the measures of convexity and compactness have been proved useful in many music related problems, they cannot contribute to the question of small group segmentation of monophonic melodies[4]. It is possible that the measures of convexity and compactness could be useful in segmentation and similarity for longer polyphonic melodies, but this could not be tested in this project since it would require a substantial research effort to extend of the ISSM, which was deemed to be outside the scope and time frame of this project.

References

Bod, R. (2002). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research 31*(1), 27–37.

Buteau, C. (2005). Automatic motivic analysis including melodic similarity for different contour cardinalities: application to Schumann's of foreign lands and people. In *Proceedings of the International Computer Music Conference*, Barcelona.

Honingh, A. K. (2006a). Convexity and compactness as models for the preferred intonation of chords. In *Proceedings of the Ninth International Conference on Music Perception and Cognition (ICMPC 9)*, Bologna, 22-26 August.

Honingh, A. K. (2006b). *The Origin and Well-Formedness of Tonal Pitch Structures*. Ph. D. thesis, University of Amsterdam, The Netherlands.

Honingh, A. K. (2006c). Pitch spelling using compactness. In *Proceedings of the Ninth International Conference on Music Perception and Cognition (ICMPC 9)*, Bologna, 22-26 August.

Honingh, A. K. (2007). Automatic modulation finding using convex sets of notes. In *Proceedings of Mathematics and Computation in Music (MCM2007)*, Berlin, Germany May 18-20.

Honingh, A. K. (2008a). Compactness in the euler lattice: A parsimonious pitch spelling algorithm. *Musicae Scientiae*. to appear.

Honingh, A. K. (2008b, April). A geometrical approach to find the preferred intonation of chords. Technical Report TR/2008/DOC/02, Department of Computing, City University London.

Honingh, A. K. and R. Bod (2005). Convexity and the well-formedness of musical objects. *Journal of New Music Research 34*(3), 293–303.

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

Schaffrath, H. (1995). The Essen folksong collection in the humdrum kern format. Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Weyde, T. (2002). Integrating segmentation and similarity in melodic analysis. In *Proceedings of the International Conference on Music Perception and Cognition 2002 (ICMPC)*, pp. 240–243.

Weyde, T. and K. Dalinghaus (2003, April-June). Design and optimization of neuro-fuzzy-based recognition of musical rhythm patterns. *International Journal of Smart Engineering System Design 5*(2), 67–79.

[4] At first sight, it seems to be the case that with these simple melodies, rules like the Gestalt principles are much more indicative of phrase boundaries than the rules of convexity and compactness (see for example Bod (2002))