Daniel Müllensiefen
Geraint Wiggins

Polynomial functions as a representation of melodic phrase contour

Abstract

The present paper explores a novel way of characterising the contour of melodic phrases. Melodic contour is represented by a curve that can be derived from fitting a polynomial equation to the pitch values given the note onsets. To represent contour numerically, we consider the coefficients of the polynomial equation as well as the time limits of the melodic phrase. After a brief review of different theoretical, didactic, analytic, and computational approaches to melodic contour, a detailed step-by-step description is provided of how polynomials can be fitted to melodic phrases. In two application examples, it is demonstrated how polynomial contour can be used as a basis for the computational processing of melodic information. The first application estimates the frequency of occurrence or prevalence of a phrase contour: a probabilistic model is constructed based on a large sample of phrases from popular melodies using a standard density estimation procedure to obtain an indicator of contour occurrence frequency. The second application is a similarity measure that exploits polynomial curves graphically and is based on Tversky's (1977) ratio model of similarity. Further applications of the approach as well as quantitative tests of the existing ones are discussed as options for future work.

1 Introduction

1.1 Background

Melodic contour is often regarded as one of the most important features in the analysis and composition of melodic music, i.e. music that is mainly conceived as consisting of one or several horizontal lines or voices. Throughout history theorists have stressed the importance of the melodic line for the construction of a *good melody*. Examples include Ernst Toch, who, in his *Melodienlehre* (1923), defines a melody as consisting of a line of pitches ("Tonhöhenlinie") and a rhythm, and Knud Jeppesen (1935), who mentions in his counterpoint text book the importance of a good balance between melodic skips and related step-wise motion, as well as the special role that contour extrema have in the construction of melodic lines. Rules to relate contour extrema to each other in a musically meaningful fashion are also given by Paul Hindemith in his *Unterweisung im Tonsatz* Bd. I (1940). The notion of melodic contour features even more prominently in Walter Piston's book on counter point (1950) where he introduces "the melodic curve" as a concept for describing and structuring melodies in the very first chapter. A recent survey of 24 historical melody composition treatises from the 1700s to the present (Winkelhaus, 2004) identifies melodic contour as one of the basic global characteristics according to which melodies can be defined, classified, and constructed. Popular song and melody writing books concur, and employ the concept of "melodic contour"

(Perricone, 2000; Bradford, 2005) or "melodic shape" (Kachulis, 2003) for their respective instruction methods.

Music analysts, too, make frequent use of the concept of contour when they describe and interrelate sections of melodies or melodic phrases. Meyer (1956) demonstrates the basic perceptual Gestalt laws of *good continuation, completion* and *closure* by the "motion" of a melodic line within a number of examples. Kunst uses the "melodic curve" as a structural descriptor within his logical approach to analysis (1978, e.g. 113). Grabner (1959) classifies the possibilities for the melodic motion ("melodischerVerlauf") of instrumental melodies and also proposes principles of alteration and variation of melodic motives and phrases. Kühn, in his *Formenlehre der Musik* (1987), points out that alterations of rhythmic and melodic motion ("Tonhöhenverlauf (Diasthematik)") are often balanced, especially in compositions from the classical era, to maintain the understandability or recognisability of the melodic unit. The same point is made by Rosen (1971) in several of his in-depth analyses of compositions in the classical style. De la Motte in his workbook on melody (1993) uses the term melodic arch ("melodischerBogen") to denote the gross motion of pitch over time. Like many of the aforementioned authors he stresses the balance between steps and skips, the role of the melodic range, and the direction of contours of subsequent segments of a melody.

From the very beginning of the discipline of music cognition, in the 1970s, and probably inspired by extensive analytic and compositional literature, researchers have investigated the question of how of melodies are perceived by a listener and what role melodic contour plays in perception. One of the most active and influential researchers has been W. Jay Dowling, who has demonstrated in several studies the importance of contour for melodic memory, particularly if melodic phrases are transposed, not presented in context, or with other musical parameters that carry little information (e.g. Dowling & Fujitani, 1971; Dowling & Bartlett, 1981; Dowling & Harwood, 1986; Dowling et al., 1995). Dowling (1978) published a very influential theory of melodic memory, according to which, melodic contour and scale information are the main informational dimensions processed in memory and are probably sufficient to reconstruct the full melody. In addition, several psychological studies by other authors have confirmed the important role that contour plays in the reconstruction of other melodic parameters in memory (e.g., Idson & Massaro, 1978; Eiting, 1984; Cutietta& Booth, 1996). The requirements for contour to act as an effective cognitive representation are that the contour of a phrase has to be simple (few changes of direction, symmetrical structure) and that the task does not allow for direct encoding of absolute pitch structure (i.e., the task involves transposition; see e.g. Idson & Massaro, 1976; Cuddy & Lyons, 1981; Taylor & Pembrook, 1984; Edworthy, 1985; Boltz & Jones, 1986). One outcome from a study by Dowling and colleagues from 1995 is the conclusion that the encoding of contour as a conscious process is of great importance when listeners are aware of the memory task and try consciously to reconstruct other parameters in memory as well. In contrast, the encoding of interval and scale information appears to be a rather automatic process that does not require voluntary attention. Taken together with evidence from infant studies (e.g., Trehub et al., 1984), these experimental findings demonstrate that melodic contour is a very important concept for the conscious perception of melodies.

Therefore, given the great and obvious usefulness of contour as an abstracted feature, it is no surprise that there have been several attempts to define melodic contour formally in order to enable the algorithmic computation of a contour representation for any given melodic phrase.

One of the most simple and reductive definitions and implementations of melodic contour was proposed by David Huron (1996). His definition of contour is based on the pitch height of the first and last notes of a melodic phrase and the mean average pitch[1] of all notes in between. Huron describes phrase contour in terms of the relationships between the first note and the average note and the average note and the last note, expressed simply as higher, lower or equal, with no information about the size of the interval. From the combination of these two free parameters with three different possible values, Huron defines nine distinct contour classes to which any melodic phrase can be assigned. The classes are named according to the two-dimensional visual shapes the three notes describe: *convex, concave, ascending, descending, horizontal, horizontal-descending, horizontal-ascending, ascending-horizontal,* and *descending-horizontal*. A classification into effectively the same nine contour classes is used by Galvin et al. (2007) to test the auditory discrimination abilities of cochlear implant users. In that study, which makes no reference to Huron's work, the nine contour classes are employed for not analytical purposes but to generate test stimuli, with absolute frequency level, interval size, and contour being the three independent dimensions of the artificial test melodies.

The advantages of Huron's contour definition are that a) any melodic phrase can simply be assigned to one out of nine contour categories, b) the categories are easily distinguishable and correspond in part to contour description used by theorists (see, e.g., the close correspondence to the five categories defined by Perricone, 2000), and c) the contour classes are very easy and quick to compute. Comparing melodic contours is very straightforward since the only existing relation between contour classes is identity or difference.

The drawbacks of Huron's contour definition are that it reflects neither the rhythmic dimension nor the length of the melodic phrase. Furthermore, it does not represent any difference between melodies with very wide or very narrow pitch range—an attribute that is considered important by many theorists. For example, both the simple three-note motif c' d' c' and a long melody with many interval skips and extensive scalar movements, but which happens to start and end on low notes, would be simplistically allocated to the same *convex* contour.

While Huron's definition of melodic contour is probably the most reductive one, the so-called *step contour* as defined by Steinbeck (1982), Juhasz (2000), Eerola and Toiviainen (2004) and others is the most literal contour definition. Here, each note is represented by its onset and duration and its pitch height. The contour of a melody can therefore be represented graphically as a step curve where the x-axis represents time and the y-axis is pitch. For the duration of every note starting at its onset on the x-axis, the curve denotes a constant pitch height value (y) until the onset of the next notes. In order to make the step curve representation invariant with respect to absolute pitch height and absolute time, only pitch intervals and inter-onset time intervals are used. The time and pitch height dimension can be normalised to start at 0 of the coordinate system (i.e., the

---

[1] The average note is, of course, only a notional note, and may not occur in the melody.

first onset starts at time 0 and either the first or the lowest pitch value is assigned the value 0 as well).

The advantage of the step contour representation lies in its precision: every contour movement, even by the shortest and most insignificant note, is represented. The disadvantage is that it does not summarise or abstract the melodic contour at all and the space of melodic contours is as large as the space of melodies. So this representation does not usefully assist comparison between, or classification of, melodic contours.

Between these two extremes of strong reduction (Huron contour) and information-preserving transformation (step contour) many alternative ways of representing melodic contour have been proposed in the literature. There is not space here to review all of them here in depth. Nonetheless, it is worth at least mentioning some of the elegant ideas that have been suggested in order to define and implement melodic contour. The definitions of contour differ according to the purpose for which the contour information is required.

Interpolation contour as proposed by Steinbeck (1982) or Zhou & Kankanhalli (2003) can be regarded as a sophistication of the step curve representation. Like in these two publications, this representation is often employed as a first transformation and reduction stage in melodic similarity measures. The basic idea is to join the turning points or extrema of a melody by straight lines. Minor contour changes such as those generated by change notes, appoggiature, and other ornaments should be excluded as they are not considered important for the overall contour motion; rules for doing so differ between authors (Müllensiefen & Frieler, 2004). To implement this representation of contour, the sequence of length and gradient values have been used, and this usually results in an effective reduction of the melodic data. Alternatively, if melodic contour is used just as a transformation, without the intention of reducing or summarising melodic data, then taking pitch height values at onset points or at regular intervals from the interpolation line is another option (e.g., Steinbeck, 1982).

Schmuckler (1999) compares tonal and dodecaphonic melodies with an algorithmic similarity measure which uses melodic contour as its core representation. He models melodic contour as a smooth up and downward motion, which is approximated by overlaid sine waves. Fourier analysis is then applied, to obtain a set of Fourier components from the ranked pitch and raw onset values. Schmuckler uses only the first six amplitude and phase coefficients to represent the contour of every melody. In contrast to Schmuckler's own experiments with dodecaphonic and artificial tonal melodies, Müllensiefen & Frieler (2004) found Schmuckler's contour-based similarity to have predictive power when compared to the similarity judgements of human experts on popular melodies.

Another family of contour representations is characterised by inclusion of all notes in the melody, but reducing the pitch representation to ordinal relations (e.g., onset note 2 higher than onset note 1) defined exclusively between adjacent notes. The simplest and widest-spread of this family is the *Parsons' code* which records whether the second note in a contiguous pair is lower or higher than, or equal to the first (symbols conventionally denoting these relations are - , + , 0, respectively). Parsons' *Directory of tunes and musical themes* (1975) lists many thousands of mainly classical melodies using this representation. However, empirical evidence from cochlear implant users (Galvin et al., 2007) suggests that most listeners are able to retain a more precise, yet still approximate, representation of interval size than merely the direction of the interval movement as encoded by the Parsons code. To this end, intervals have been classified into five and seven (Kim et al., 2000), and into nine different classes (Pauws, 2002; Müllensiefen & Frieler, 2004). In this last case, scale intervals are grouped together into discrete classes which become less discriminatory as interval size increases. For example, Müllensiefen and Frieler (2004, p.152) assign a unison to class 0, upward and downward seconds are classes +1/-1, thirds are classes +2/-2, fourths and fifths are +3/-3, and all intervals smaller or greater than that are coded as +4/-4. A strong argument in favour of this family of contour representations is that it seems to approximate human perception in as much as humans tend to be able to make very reliable judgements about the direction and approximate size of an interval. The negative aspect of the Parsons' code and related contour schemes is certainly the fact that they do not admit exact inference about pitch height relations for non-adjacent notes.

A few approaches related to the Parsons' code overcome this problem by filling a matrix between all possible pairs of notes with and indicator of interval direction (1 if note *i* is lower than note *j* or 0 otherwise) between row element *i* and column element *j*. Since these models cover ordinal relations between all note pairs of a melody, they are sometimes referred to as *combinatorial models of contour* (Shmulevich, 2004). Most of them are rooted in the theoretical models of Friedmann (1985) and Marvin and Laprade (1987), subsequently elaborated by other researchers (e.g., Quinn, 1999). The information contained in the contour relation matrix can be exploited for various purposes, one of which is to determine the similarity between two melodies of equal length by the *contour similarity index* (CSIM; see Marvin & Laprade, 1987, and Polansky, 1996). The CISM has been shown to be equivalent to Kendall's τ, a long-established ordinal measure of concordance, and is suspected not to be very robust to structural permutations of melodies (Shmulevich, 2004). While combinatorial contour models are far more comprehensive than the representation from the Parsons' code family, in that they contain information about non-adjacent notes, they may retain more information than human listeners can actually process in real time. A listener may only compare the pitch height of a given note to other notes in its temporal vicinity (e.g., within a time window which corresponds to working memory span of 2-3 seconds; see Baddeley, 1997). But, because the contour relation matrix is uniform, the interval between notes 1 and 2 is of equal importance to the one between, say, note 5 and note 55 of a very long melody. In sum, the combinatorial contour models do not reflect the linear nature of a melodic line but rather a holistic conception of contour. Also, unless some summarising method is performed on the contour relation matrix, the combinatorial models do not reduce, but inappropriately include ornaments and change notes.

1.2 Motivation

From the previous section, it emerges that there is neither a single notion of what melodic contour really is, nor a single application of it. Instead, different ways of looking at the directional movement of pitch height over time in a melody have emerged for different theoretic, didactic, analytic, and practical purposes. In the present paper, we do not aim to decide how melodic contour should be defined correctly; nor are we trying to compare different approaches or to reconcile diverging concepts of contour. Instead, we develop a new method of defining and actually computing melodic contour from a series of onset and pitch height values. This new method is ultimately designed to represent melodic phrases of a large collection of popular melodies as part of the M^{4}S project at Goldsmiths[2]. It aims to overcome some of the disadvantages of existing contour definitions and implementations. For the present purpose, some important requirements for the new contour representation are:

- The musical unit that it is primarily designed to work on is the melodic phrase—which is defined as a shorter segment of a melody that induces a certain perception of closure towards its end. Several algorithmic melody segmenters produce segmentations of melodies that correspond roughly to these assumptions (e.g. Cambouropoulos, 2001; Temperley, 2001; for an overview, see Pearce et al., 2010). The contour computation is designed to work on the output of such segmenters.
- A good balance between summarising the melodic events of a melodic phrase and discriminating between a large number of phrases should be maintained, to make melodic contour a feature which is helpful in grouping melodic phrases together and in searching and retrieving melodies from a large collection.
- The representation should be robust to minor variations in pitch and time, and do some degree of smoothing over melodic ornamentation.
- The linear evolution of a melodic phrase as generally perceived by human listeners should be emphasised, and intuitively it should be possible to relate the new contour representation to the conventional representation of a melody in staff notation.

A second aim of this paper is to demonstrate the potential of our polynomial contour representation as a means of describing and modelling a large collection of melodies, and as a basis for the construction of similarity measures between pairs of melodies.

2 A new method for computing melodic contour

The basic idea of this new method for representing the contour of melodic phrases is very simple: Imagine one connects all the notes of a melody in staff notation, like the one in Fig. 1, with a smooth line.

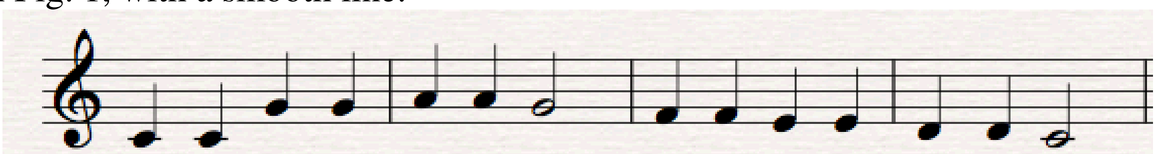

Fig. 1: First two phrases of the well-known French tune *Ah vous-dirai-je, Maman*; in English-speaking countries often sung to the lyrics of the nursery rhyme *Twinkle, twinkle little star*; in German-speaking countries it is known as the Christmas carol *Morgen kommt der Weihnachtsmann.*

Just as in this example, for many melodic phrases the result would be a curved line, maybe going up and down a few times. To represent these smooth and short curves numerically we use polynomial equations, like the one in eqn. 1, a polynomial of order 6 (meaning that the highest exponent for any term in this monovariate equation is 6).

[2] see http://www.doc.gold.ac.uk/isms/mmm/

(1) $y = 65.987 - 1.589 \cdot x + 0.589 \cdot x^2 - 0.13 \cdot x^4 + 0.024 \cdot x^5 + 0.004 \cdot x^6 - 0.001 \cdot x^7$

When computed over the range of $x \in [-4.2, 4.2]$ this equation corresponds to the graph in Fig. 2, which seems intuitively to fit the contour of the melody in Fig. 1 quite well. There is a sharp rise in the first half of the graph which approaches MIDI pitch 69 corresponding to the A4 at the beginning of bar 2. After that the curve falls down smoothly towards the end.

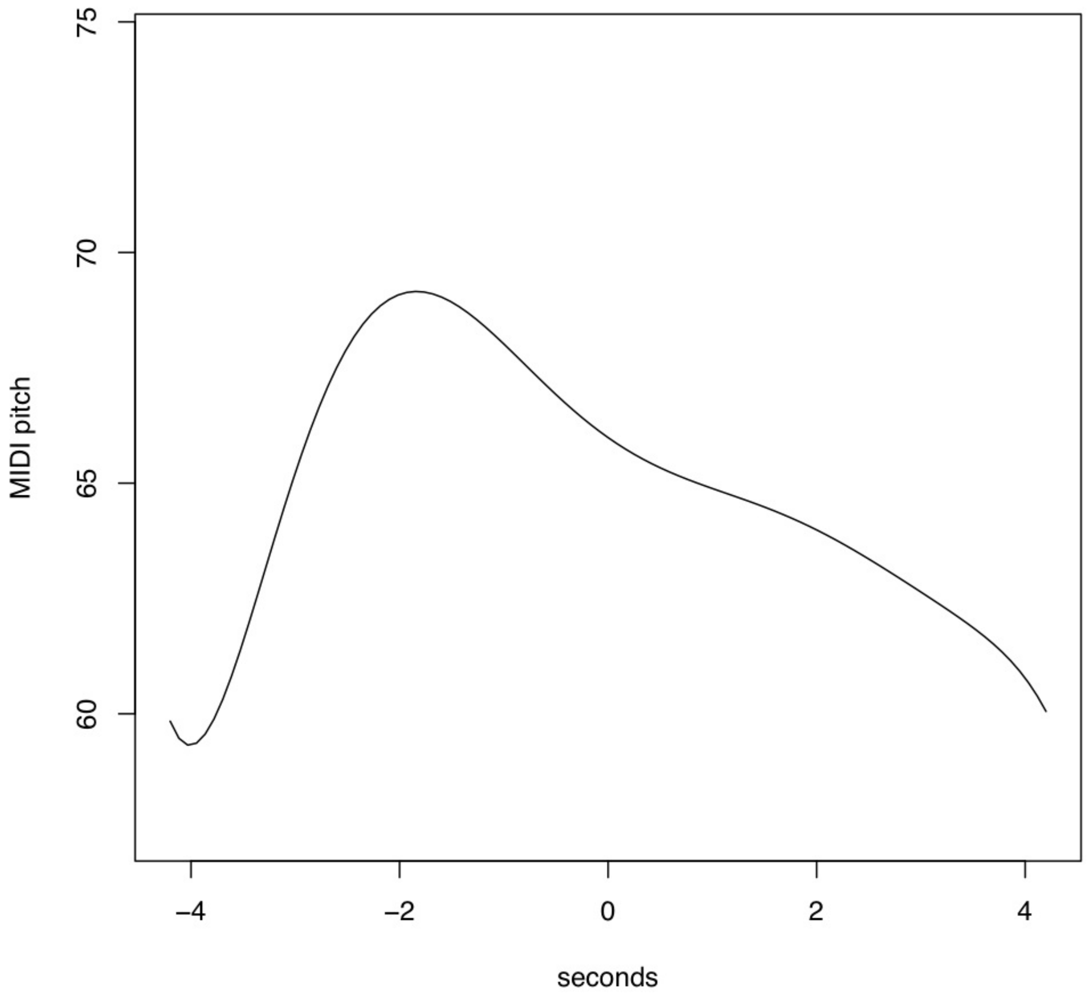

Fig. 2: Polynomial contour curve corresponding to Eq. 1 and derived from the first two phrases of *Ah vousdirai-je, Maman*.

If the second phrase (bars 3-4) of the same melody is modelled on its own we obtain the smooth descending line depicted in Fig. 3.

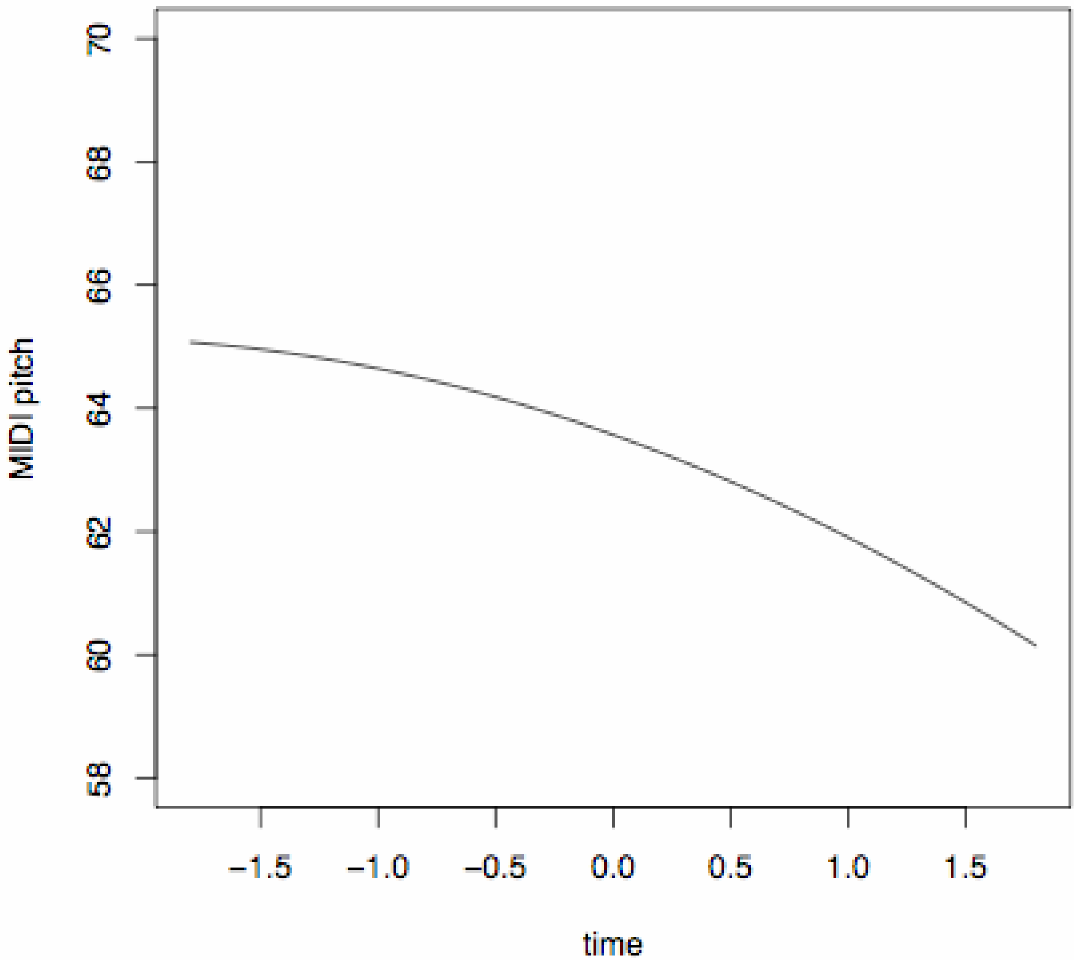

Fig. 3: Polynomial contour curve derived from the second phrase of *Ah vousdirai-je, Maman*. The corresponding polynomial equation is: $y = 63.571 - 1.369 \cdot x - 0.298 \cdot x^2$

The contour curve for the second phrase is characterised by only three components: An additive constant (63.571) that corresponds to the central pitch level, between D#4 and E4, a negative linear component ($-1.369 \cdot x$) which is responsible for the overall downward motion, and a smaller negative quadratic component ($-0.298 \cdot x^2$) which adds a bit of smoothness to the curve.

The polynomial curve for the first phrase (bars 1–2) only is given in Fig. 4.

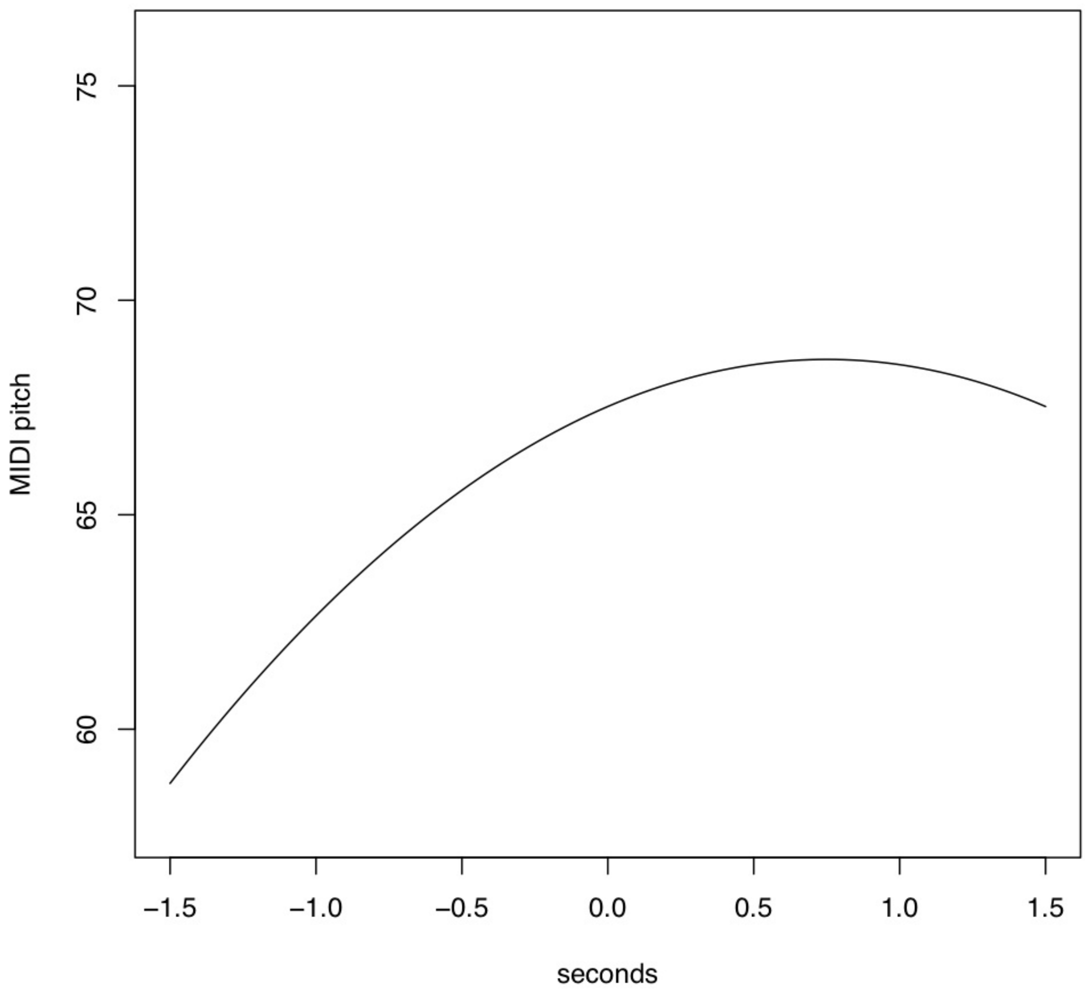

Fig. 4: Polynomial contour curve derived from the first phrase of *Ah vousdirai-je, Maman*. The corresponding polynomial equation is: $y = 67.524 + 2.929 \cdot x - 1.925 \cdot x^2$

The curve of the first phrase is dominated by a strong positive 1st-order component ($2.929 \cdot x$) which is responsible for the overall ascending trend of the curve. The comparatively weaker quadratic component generates the directional turn at the end of the phrase (A4 to G4).

In comparison, if we drop the last note (G4) from the first phrase, and fit a contour curve to the first six notes (C C G G A A) we obtain, as a result, an almost straight line characterised by the additive constant, an even stronger linear component ($3.429 \cdot x$) and a very small quadratic component that models the lower steepness of the contour going only a major second up from G4 to A4 (Fig. 5).

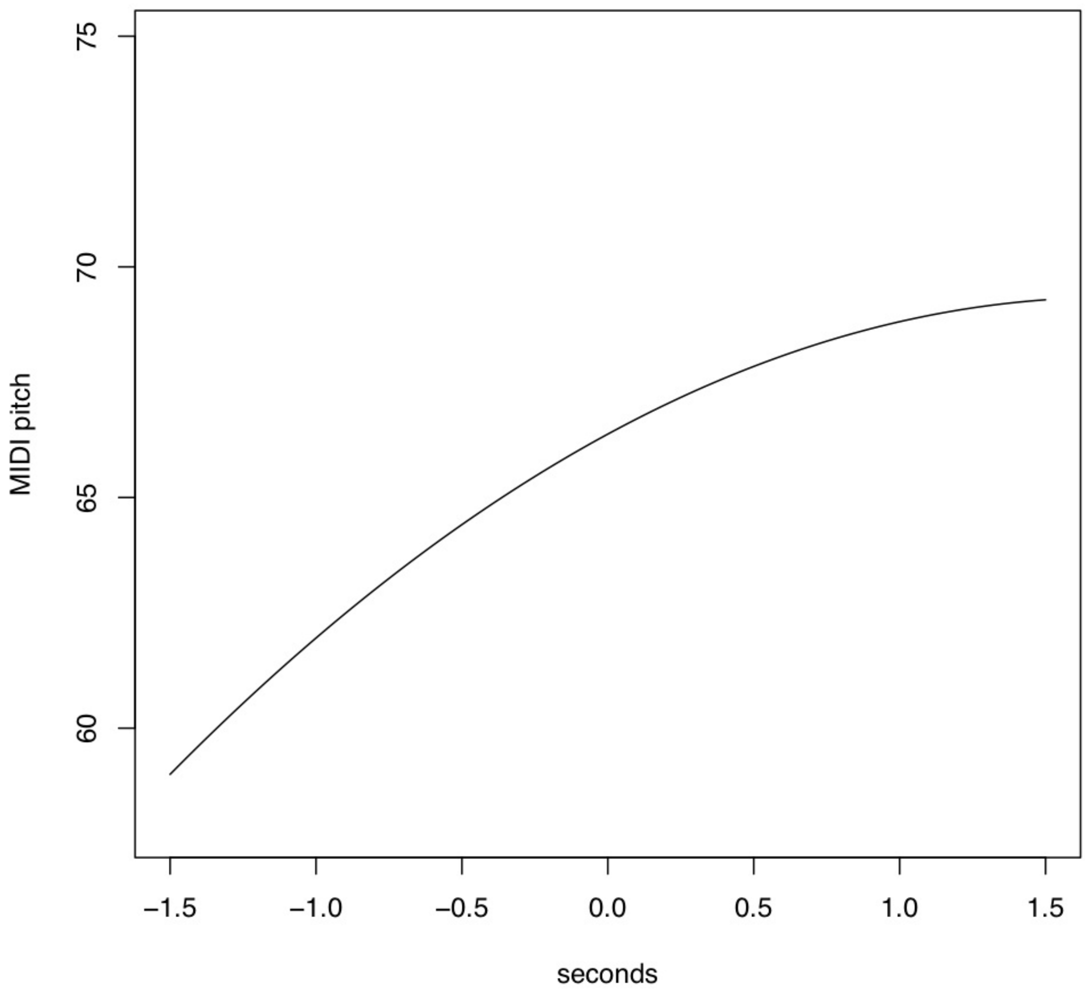

Fig. 5: Polynomial contour curve derived from the incomplete first phrase (first 6 notes) of *Ah vousdirai-je, Maman*. The corresponding polynomial equation is: $y = 64.375 + 3.429 \cdot x - 0.992 \cdot x^2$

This simple example makes the central points about the polynomial contour representation very clear: It is possible to represent a melodic phrase by a polynomial. The polynomial curve aims at connecting the pitches of the phrase in a smooth way and thus captures the up- and downward motion of the phrase over time. The coefficients of the polynomial equation can be interpreted as a numerical representation of the melodic phrase.

Polynomials have some advantages that make them suitable for describing melodic contours. They are elegant in the sense that they only make use of the three most basic arithmetic operations (add, subtract, multiply). Simple shapes (lines, parabolas) correspond to simple polynomial equations (i.e., few and lower-order coefficients). Simple shapes can be added to form more complex ones. Polynomials are good for describing symmetrical behaviour (i.e. up and downward motion). They are simplest when describing motions with only a few directional changes, which makes them suitable for most short melodic phrases.

3 How to compute polynomial contour

The idea of representing the contour of a melody by a polynomial equation whose curve is fitted to the notes of a melody was first proposed by Steinbeck (1982). The main purpose for Steinbeck was to find a compact representation that could be used for measuring the similarity between folk melodies. Although his description of how to derive a polynomial contour representation from melodic data is quite detailed (ibid., pp. 116-121) he did not make use of the polynomial contour representation for the clustering of folk melodies that he presents later in his book but resorted to aggregate features. This is probably due to the computational limitations regarding speed and arithmetic accuracy at his disposal in the early 1980s.

Frieler et al. (in press) took up Steinbeck's original idea and represented a large collection of melodic phrases from highly commercial western pop songs by coefficients of fitted polynomials. Like Steinbeck's, the purpose of their investigation was to cluster melodic phrases into a small number of discrete classes. In their results, Frieler et al. compare clusterings performed on the basis of polynomial contour to the contour classification as generated by Huron's contour representation and find general agreement between the two methods. They stress that one of the additional benefits of the clustering based on polynomial contour is the fact that, unlike Huron's method, information about the relation between an individual melodic phrase and the centre of the cluster it belongs to is retained. Thus, a distinction can be made between phrases that are very typical for a class of phrases and those that can be considered as rather atypical or outliers. The approach we present here for obtaining and working with a polynomial contour representation differs in a number of ways from Steinbeck's original idea and from the empirical adaptation as engineered by Frieler and collaborators. We will highlight these differences of the present implementation from its predecessors below, where appropriate.

The computation of polynomial contour is carried out in four consecutive steps:

1. Representation and selection of melodic data: segmentation and phrase length limits

A melody m is represented as a series of pairs of pitch and onset time value (p_i,t_i) for each note n_i of the melody m. In principle, the units for these values could be represented with reference to any suitable measurement system. In practice, we use MIDI pitch since the main collection of popular music we are currently investigating has MIDI as its source format. The decision to use absolute timing information in milliseconds instead of metrical time as measured in fractions of a beat is rooted in the assumption that from a certain point on the tempo of a melodic phrase acts on the perception of melodic contour and the variation in tempo in our collection of popular songs is too large (range 6bpm to 303bpm) to be ignored. Our definition of contour is supposed to work on melodic phrases as units. However, just as with melodic contour itself, there exists a considerable debate among music theorists and psychologists about what constitutes a melodic phrase and how a continuous melodic line should be segmented or grouped into shorter meaningful units (e.g., Temperley, 2001; Pearce, Müllensiefen, & Wiggins, 2010). Several algorithms have been proposed for segmenting melodies into consecutive phrases and a few comparative studies have tested these segmentation algorithms with respect to their cognitive validity (e.g., Thom et al., 2002; Müllensiefen et al, 2008; de

Nooijer et al., 2008). We chose Grouper, David Temperley's rule-based algorithm (Temperley, 2001), as a segmenter since it has performed well in comparative studies using different melody collections and because it can handle unquantised data by using a built-in quantisation mechanism. For each melody *m*, Grouper produces melodic phrase boundaries according to which the notes of a melody can be grouped together into a number of phrases φ_i. The number of notes naturally varies between phrases. In order to obtain a homogeneous collection of melodic phrases for the present study, we only consider phrases with a length within the 95-percentile around the mean phrase length (9.28 notes) from the Essen folk song collection (Schaffrath, 1995)[3]. To put it differently, we discard all phrases that have a number of notes significantly different from the mean phrase length of a large collection of popular melodies with manually annotated phrase segmentations. The 95-percentile corresponds with limits of $4 < N < 16$ for the number of notes of a melodic phrase. Also, as we are using absolute timing information, it is necessary to set a limit to the phrase duration as measured in milliseconds. 95% of the 399,672 melodic phrases in the M^4S database, resulting from the segmentation by Grouper and between 5 and 15 notes long, have a duration between 0.98 and 6.85 seconds. We limit ourselves to the analysis of melodic phrases within these absolute time limits.

Unlike Steinbeck and Frieler et al., we do not standardise the duration of the phrases to unit time. This would eliminate effects of phrase lengths in terms of absolute time or number of notes and tempo. However, we do assume that the natural length of a melodic phrase makes a difference to contour perception, which we would like to be reflected in the polynomial model. Frieler et al. are also aware of the perceptual importance of phrase length (as measured in the number of notes in a phrase) and retain this information in a variable separate from the polynomial model.

2. Transformation of melodic data: Centring around the origin

To exploit the symmetrical nature of the polynomial function (reflectively symmetrical with respect to the y-axis for even exponents, and rotationally symmetrical around the origin for odd exponents), we centre each melodic phrase around the origin on the time scale. This is achieved by shifting all onset values to the left on the time scale according to equation 2:

(2) $$t'_i = t_i - \left(t_1 + \frac{t_n - t_1}{2}\right)$$

Under the assumption that many melodic phrases have a more or less symmetrical structure, centring around the origin reduces the number of parameters to be estimated and is therefore a clear improvement which has not been considered by Steinbeck nor Frieler et al.

---

[3] The grouping into phrases in the Essen folk song collection was carried out manually by expert musicologists and folk song collectors.

3. Fitting the full model

As we are striving towards a summary representation of contour, we set $\lfloor n/2 \rfloor + 1$ for a phrase φ with n notes as an upper limit for the number of free parameters to estimate. For a given phrase φ with n notes the full polynomial model is therefore defined as

$$p = c_0 + c_1 t + c_2 t^2 + \ldots + c_m t^m \qquad (3)$$

where $m = \lfloor n/2 \rfloor$. To obtain the parameters, c_i, we use least squares regression and treat the exponential transformations, t, t^2 … $t^{n/2,}$ of the onset variable t as predictors and pitch height p as the response variable. Only the actual time and pitch values (t_i, p_i) of the melodic phrase are used in the regression procedure. The result is a numerical value for each of the parameters c_0, .., $c_{\lfloor n/2 \rfloor}$.

4. Variable selection

Least squares regression generally overfits the data of the response variable—that is to say, the regression model is likely not to be robust to minor variations in the data. In our case this could mean that melodic phrases with similar contours would be represented by hugely different sets of polynomial parameters. Therefore, we apply Bayes' Information Criterion (BIC) as a standard variable selection procedure that balances the fit to the response variable against the complexity (number of terms) of the model. The stepwise search through the model space is done backwards (starting from the full model) and forwards (starting from the c_0 parameter alone), and, at each step, predictor variables that lead to a better BIC value are included or excluded from the model. Most of the time, this variable selection process results in a final contour model where the number of predictor variables necessary to predict p is considerably smaller than $n/2 + 1$. Since all predictor variables are considered for inclusion or exclusion at each step the resulting model does not necessarily consist exclusively of exponential transformations of t with neighbouring exponents (e.g., $p = c_0 + c_1 t + c_2 t^2 + c_3 t^3 + c_4 t^4$). Instead, in many instances just a single transformation of t (e.g., $p = c_0 + c_3 t^3$) is selected, or predictor variables with non-adjacent exponents form the model (e.g., $p = c_0 + c_1 t + c_3 t^3 + c_6 t^6$).

The variable selection process ensures that the polynomial model is not over-fitted and this is a major advantage over the procedures suggested by Steinbeck and Frieler et al. who respectively determined the upper limit for the order of the polynomial with regard to the fix point precision of the computer used and the use of a fixed order (e.g., 2 or 4) for the polynomials fitted to all phrases of a collection regardless of their melodic structure.

The result of the variable selection process is a set of polynomial coefficients that are taken to represent the melodic phrase that has been modelled. This includes the coefficients that have a numerical value $\neq 0$ as well as the information about which variables (exponential transforms of t) were not selected for the contour model and therefore have a value of 0. Since the length limit of a phrase is 15 notes and the complexity limit is $\lfloor 15/2 \rfloor = 7$, we represent each melodic phrase as an 8-dimensional vector

$$\langle c_0, c_1, c_2, c_3, c_4, c_5, c_6, c_7 \rangle$$

where the c_i are the coefficients of the polynomial terms[4]. In addition to the polynomial coefficients, the shifted limits on the time axis which correspond to the onset of the first and last note of the phrase are also part of contour representation.

4 Applications

This new representation allows several interesting analytical and computational applications.

4.1 Density estimation from a probabilistic model

Like Huron (1996), we are interested in the distribution of melodic phrase contours in large collections of melodies. Knowing what very common and very uncommon contours look like can inform, for example, the automatic detection of unusual phrases in a melody or the construction of cognitive models of melodic processing which are based on the assumptions that frequency information, acquired by statistical learning, is an important aspect of many cognitive mechanisms. Also, knowledge about the distribution of melodic contours for different musical repertoires can help identify the structural characteristics of different styles, and supports theorising about musical universals and the compositional and performative constraints for vocal melodies in general (see, e.g., Huron, 2006).

To model a collection of polynomial phrase contour representations, we first standardise the coefficients for each phrase to make them comparable. We use the *z-transformation*, which normalises the raw coefficients c_i of each of the 376,423 phrases[5] from our pop song database, according to the mean and standard deviation of the coefficients of each phrase, as shown in Eqn. 4; s_c is the standard deviation. Note that the standardisation is done separately for each phrase.

$$(4) \qquad c_i' = \frac{c_i - \bar{c}}{s_c}$$

The z-transformation reduces differences between contours with respect to the interval range and temporal span. For example, arch-like contours ranging for two octaves or for only a fifth will have very similar sets of coefficients. Similarly, arch-like contours spanning five crotchets (e.g., 1 second) or 17 crotchets (e.g., 4s) having similar

[4] However in practice, when applied to a collection of 376,423 melodic phrases from popular melodies the variable corresponding to the polynomial component of order 7 was never selected. Thus, for modelling our collection of popular melodies we represent polynomial contour only as a 7-dimensional vector.

[5] The melodies were all taken from the vocal line of the song. In general, the MIDI transcriptions of the pop songs are very accurate but due to their commercial nature they sometimes contain transposed melodic segments in unsingable registers, or instrumental melodies which were saved to the vocal track of the MIDI file. Since our aim is here the characterisation of phrases from vocal melodies, we excluded all phrases with a central pitch (1st coefficient) smaller than MIDI pitch 36 or larger than 96 (most of which are probably computer generated and not actually sung phrases), as well as all phrases with coefficients outside the 99.9% percentile for any coefficient. As a result, from the original 379,703 phrases 3,280 were excluded.

coefficient values after z-transformation. The transformed coefficients do not reflect the absolute dimensions of the original curve, but information about the relative sizes within the set of coefficients is retained. Therefore, we regard the z-transformation to be a sensible step because we are interested, at this stage, in the coarse outline of the melodic contour, and not in individual details of length, tempo, duration, or interval range. Also, for this comparative contour modelling, we are not interested in the absolute pitch of the phrase, and so we disregard the additive constant, c_o, which is the zeroth-order component of the contour vector.

To model the distribution of melodic phrase contours, we make use of a statistical procedure known as Gaussian mixture modelling (e.g., McLachlan & Peel, 2000). This probabilistic model approximates an empirical multi-dimensional distribution by a mixture of multidimensional normal or Gaussian distributions. Each Gaussian distribution is characterised by only three parameters, its weight in the mixture, its mean, and its variance. The three parameters are estimated in such a way that the probabilities of each object (i.e., phrase contour) belonging to any of the Gaussians is maximised.

For technical reasons[6], we used a subsample of coefficients from 30,000 melodic phrases drawn at random from the overall 376,423 phrases. The model that fits this sample best is a 17-component *EEV* model with equal volume and equal shape but variable orientation of the components. The density distribution of only the first (linear) and second (quadratic) standardised coefficients is visualised in Figs. 6 and 7.

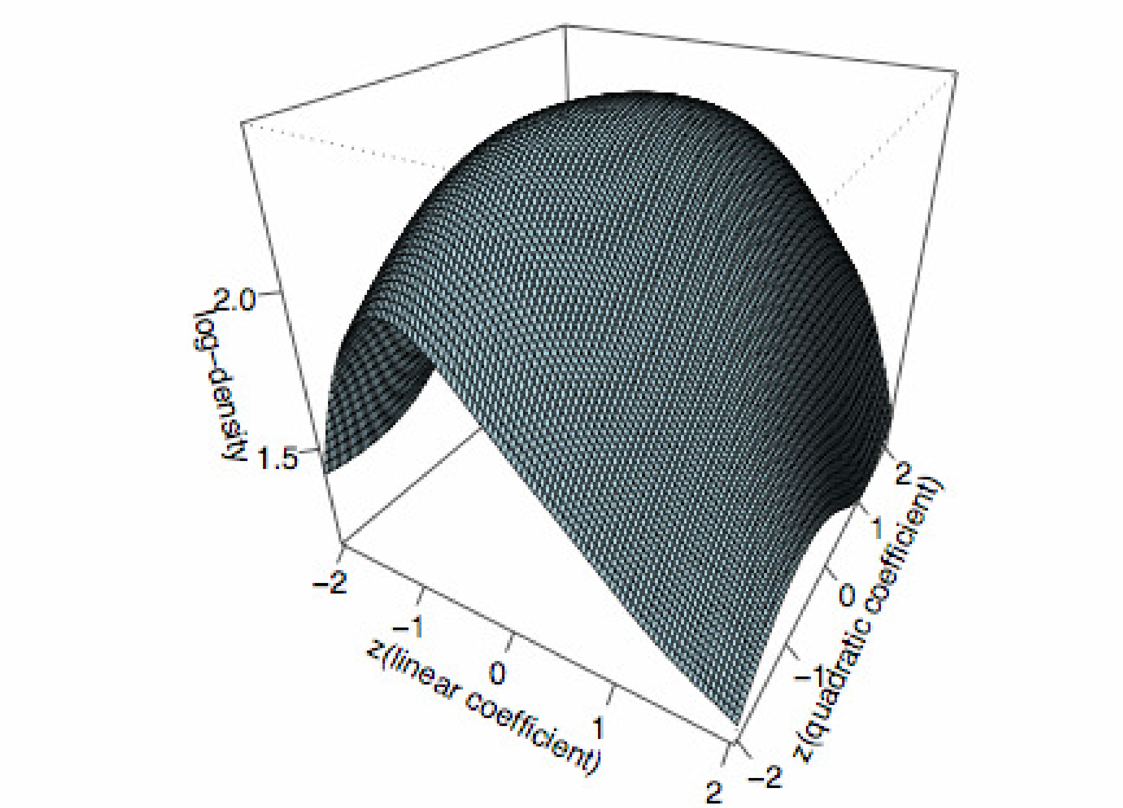

Fig. 6: Perspective plot of the log-density of the distribution of the z-transformed first and second polynomial coefficients according to a 17-component Gaussian Mixture Model. The density is marginalised over the other four coefficients.

---

[6] The algorithm we used for maximum-likelihood estimation as implemented in the library MCLUST of the statistical programming language R is very memory intensive and therefore sets an effective limit for the number of phrases to be modelled in the same model.

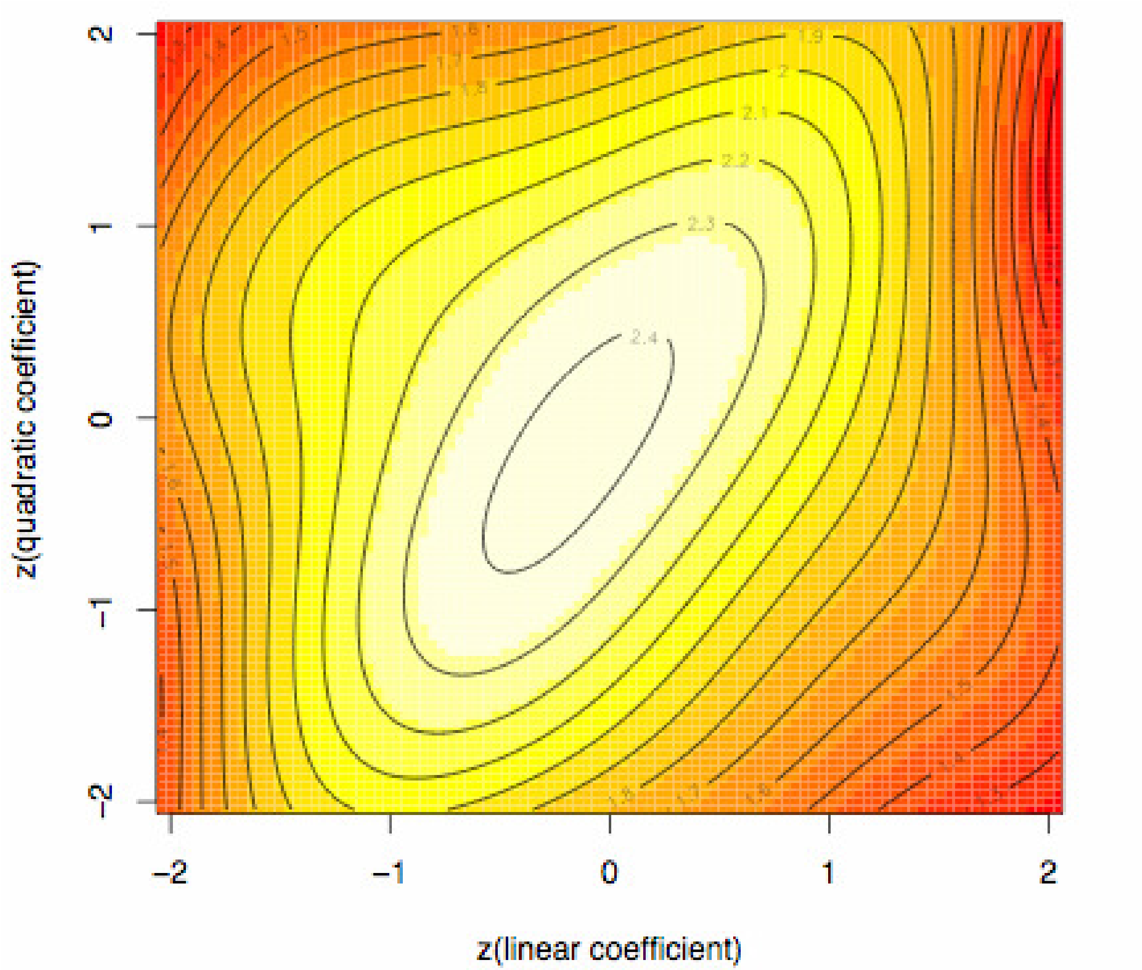

Fig. 7: Heat plot of the density of the distribution of the first and second polynomial coefficients according to a 17-component Gaussian Mixture Model. Lighter colours correspond to a higher density. The contour circles denote regions with equal density.

However, interpreting the density distributions of z-transformed coefficients is difficult and musicologically not very interesting. In that respect, an interpretation of the components of the mixture model as clusters of phrases is more rewarding. We return to this below.

With the help of the mixture model, any new melodic phrase, represented by a set of polynomial coefficients, can be assigned a density in the model space. This density can be interpreted as indicating the frequency of occurrence or prevalence of the melodic contour in the model space. As raw density numbers are difficult to interpret and depend on the dimensionality of the data and on peculiarities of the density estimation procedure, we express the prevalence in terms of the percentile into which a given phrase falls with respect to the distribution of density values. We therefore compute the distribution of densities for a new and randomly selected set of 30,000 phrases with respect to the existing model. We compute the 1-percentile boundaries of the distribution for this new set of phrases and we then can determine the percentile that corresponds to the density of the new melodic phrase contour.

For the contour of *Ah vousdirai-je Maman,* as displayed in Fig. 2, we obtain a logarithmic density value of 7.33. This corresponds to the 100th percentile of the distribution of density values of 30,000 randomly sampled melodic phrases; i.e., the contour of the melody is more common than 99% of all phrase contours. Therefore, relatively speaking, the melody displayed in Fig.1 can be considered to have a very common melodic shape.

The first and the second phrase of the same melody as displayed in Figs. 3 and 4 both have a density of approximately 5.42 and 7.49 respectively which means that they are in the 78th and 100th percentile, i.e. phrase is less common and phrase 2 is slightly more common than the contour of the first two phrases modelled together.

In comparison, we obtain a considerably lower density value for the 1st phrase of the melody of the national anthem of the USA (see Figs. 8 and 9). The density value is 3.868 which corresponds to the 41st percentile, that is, 40% of the 30,000 randomly sampled phrases have a less common contour.

Fig. 8: First phrase of the melody from the US national anthem.

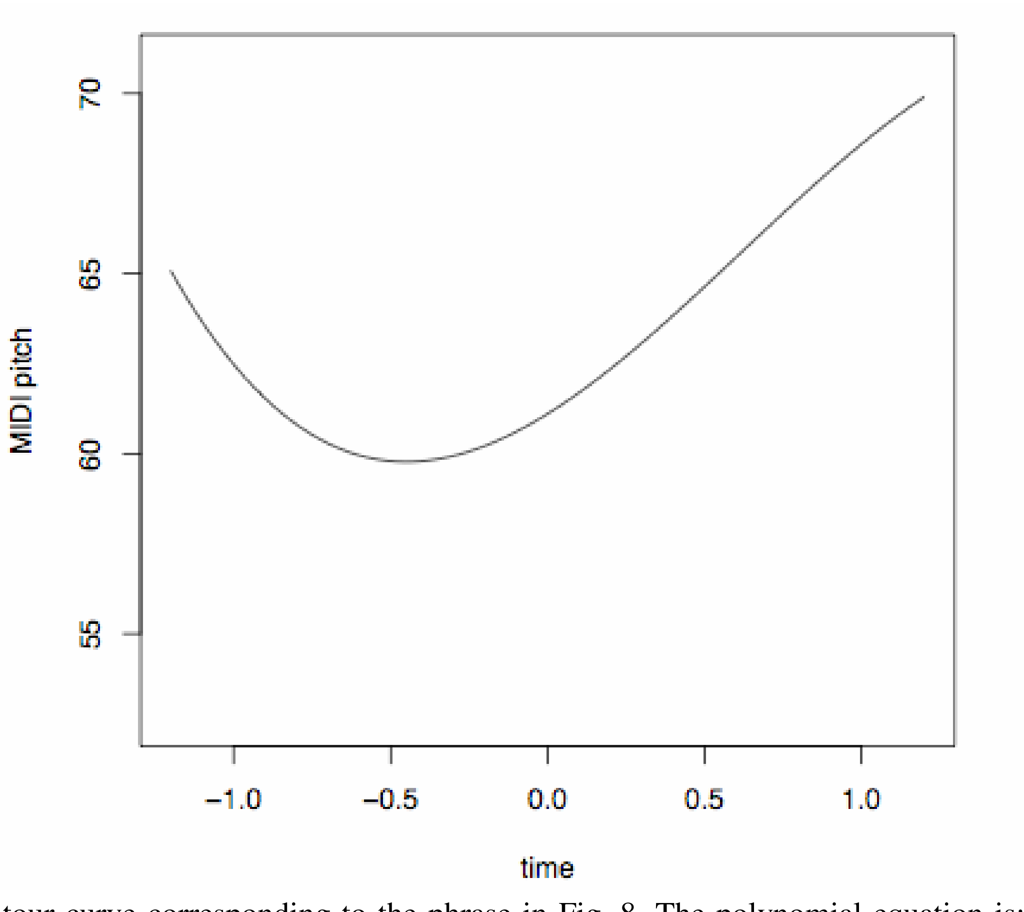

Fig. 9: Contour curve corresponding to the phrase in Fig. 8. The polynomial equation is: $y = 61.1 + 5.44 \cdot x + 4.44 \cdot x^2 - 2.38 \cdot x^3$

Apart from density estimation, Gaussian mixture models can also be exploited for clustering melodic phrases. Each multi-dimensional component can be interpreted as a separate cluster and all phrases with the highest probability of being generated from that component are considered to belong to that cluster. Due to the space limitations of this contribution, we cannot go into detail here, but a comparison to the contour-based clustering or grouping of melodic phrases as carried out by Steinbeck (1982), Huron

(1996), Juhasz (2000, 2009), or Frieler et al. (in press) would certainly be a very interesting next step.

4.2 Similarity measurement via integration

Similarity measurement has been the primary aim of most analytical and computational studies on melodic contour. We now show very briefly how the polynomial contour representation of melodic phrases can serve very straightforwardly as base representation for determining the similarity relationship between two phrases.

As an example we consider the polynomial contours of several phrases taken from Beethoven's famous theme from the 4th movement of the 9th symphony, also known as *Ode to Joy* (Fig. 10).

Fig. 10: Ode to Joy

We choose phrase 1 (bars 1, 2, and the first note in bar 3), phrase 2 (bars 3, 4), phrase 5 (bar 9), phrase 6 (bar 10), and phrase 7 (bar 11 and first 3 notes in bar 12) as example phrases for our small test set. Their polynomial contours are depicted in Figs. 11–15.

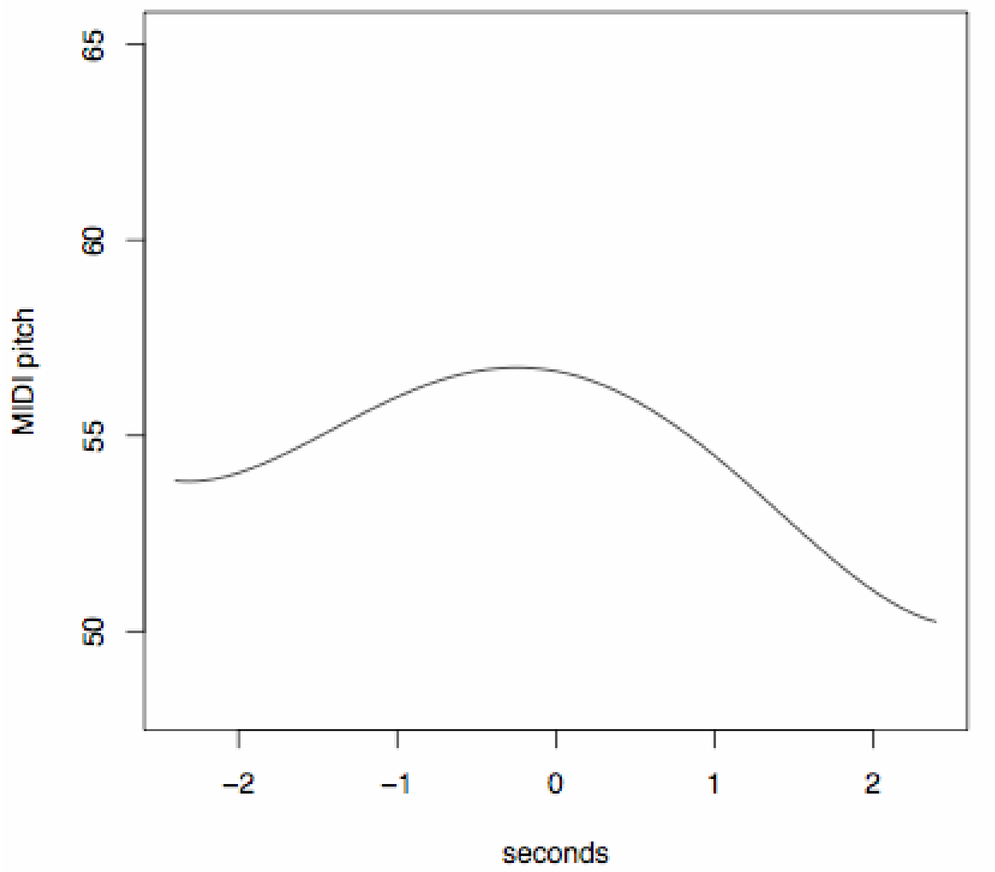

Fig. 11: Polynomial contour of *Ode to Joy*, 1st phrase

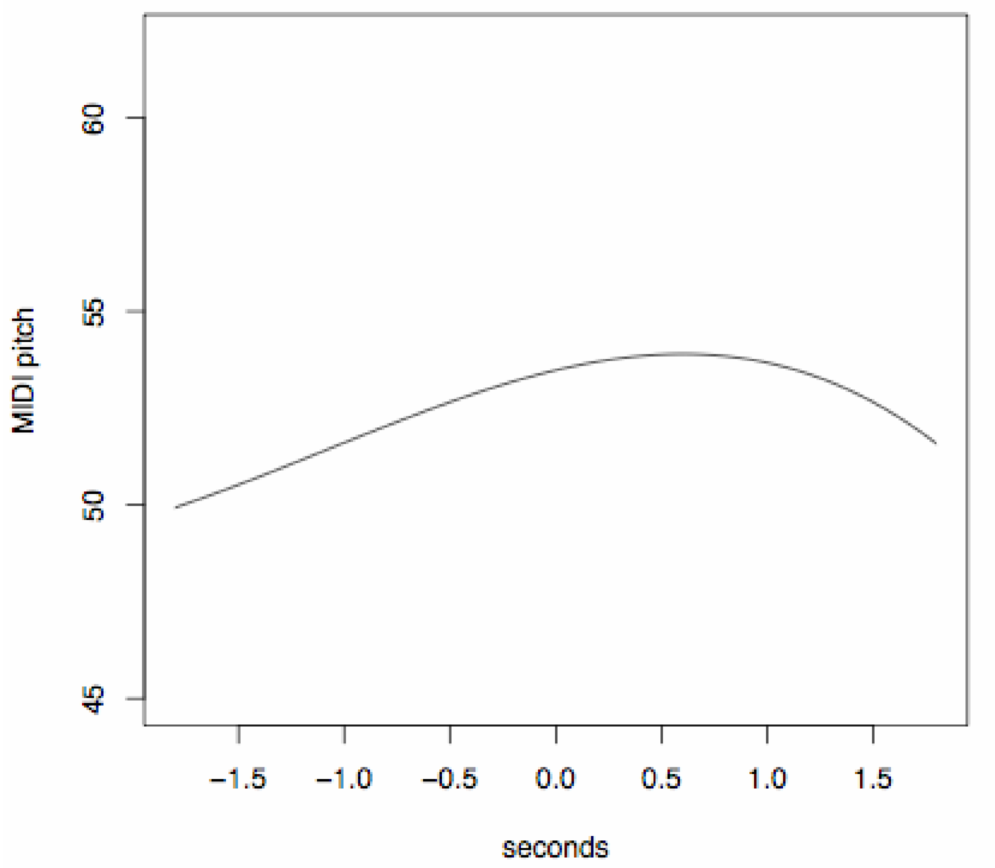

Fig. 12: Polynomial contour of *Ode to Joy*, 2nd phrase

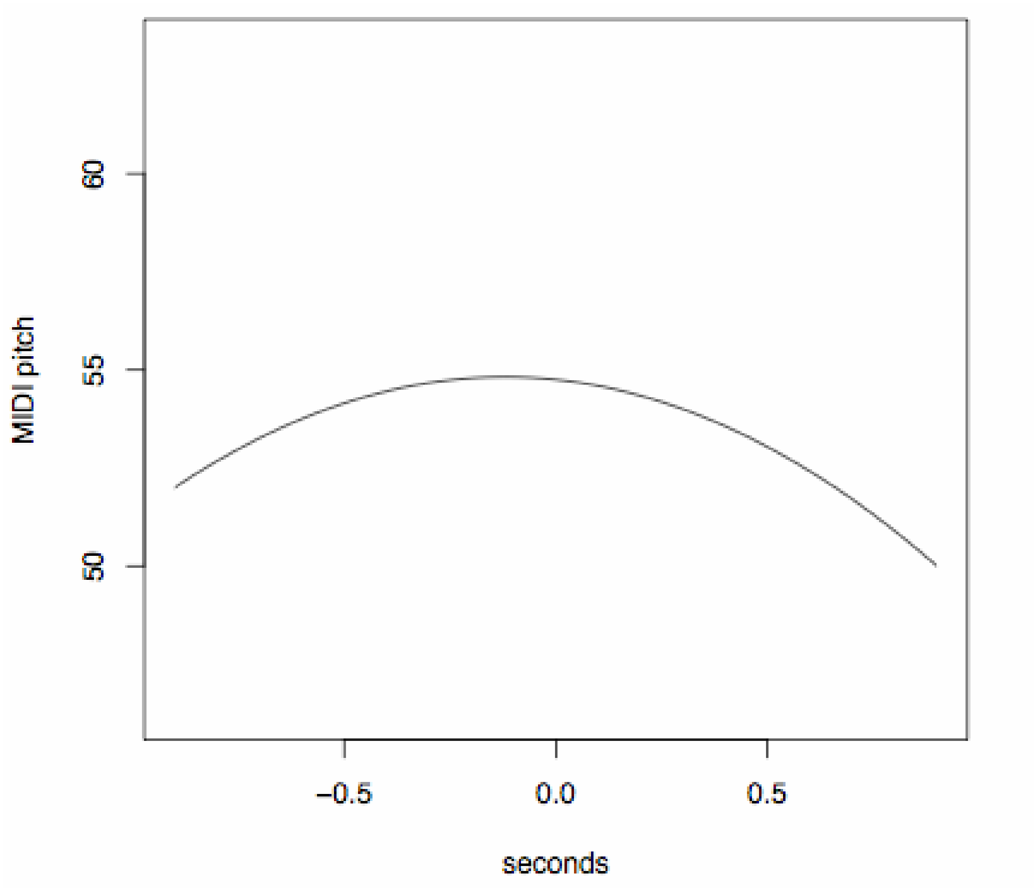

Fig. 13: Polynomial contour of *Ode to Joy*, 5th phrase

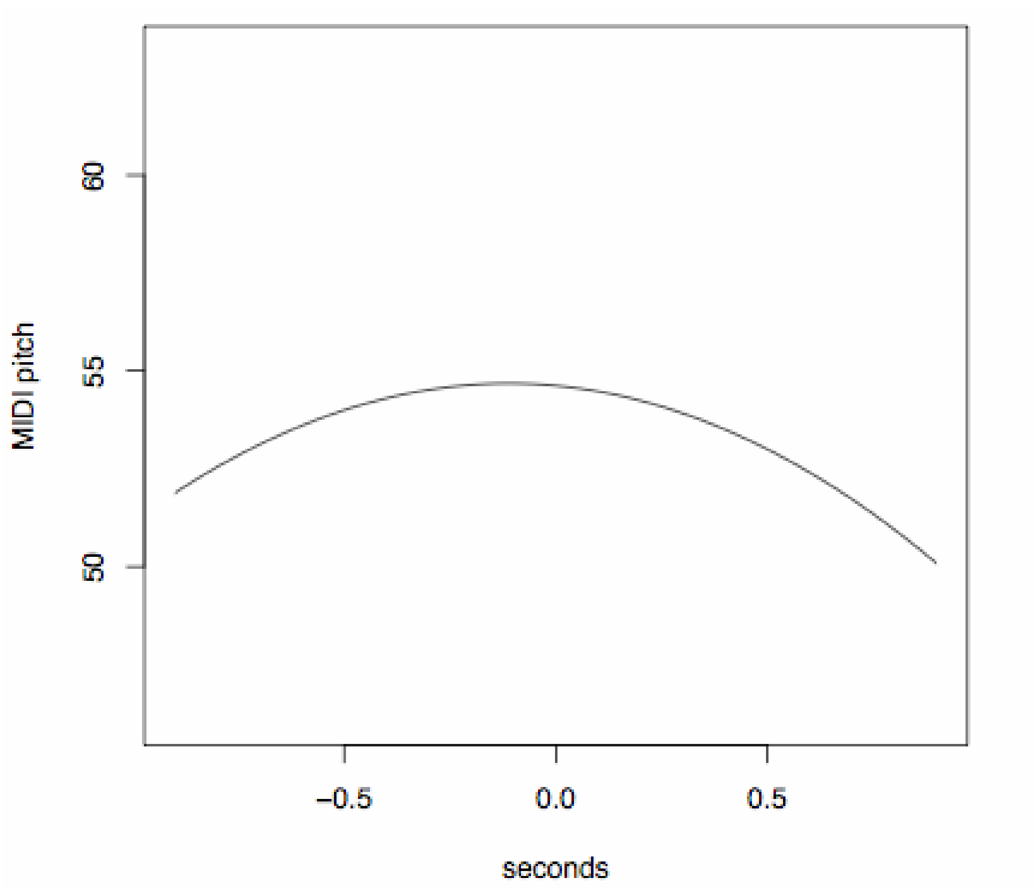

Fig. 14: Polynomial contour of *Ode to Joy*, 6th phrase

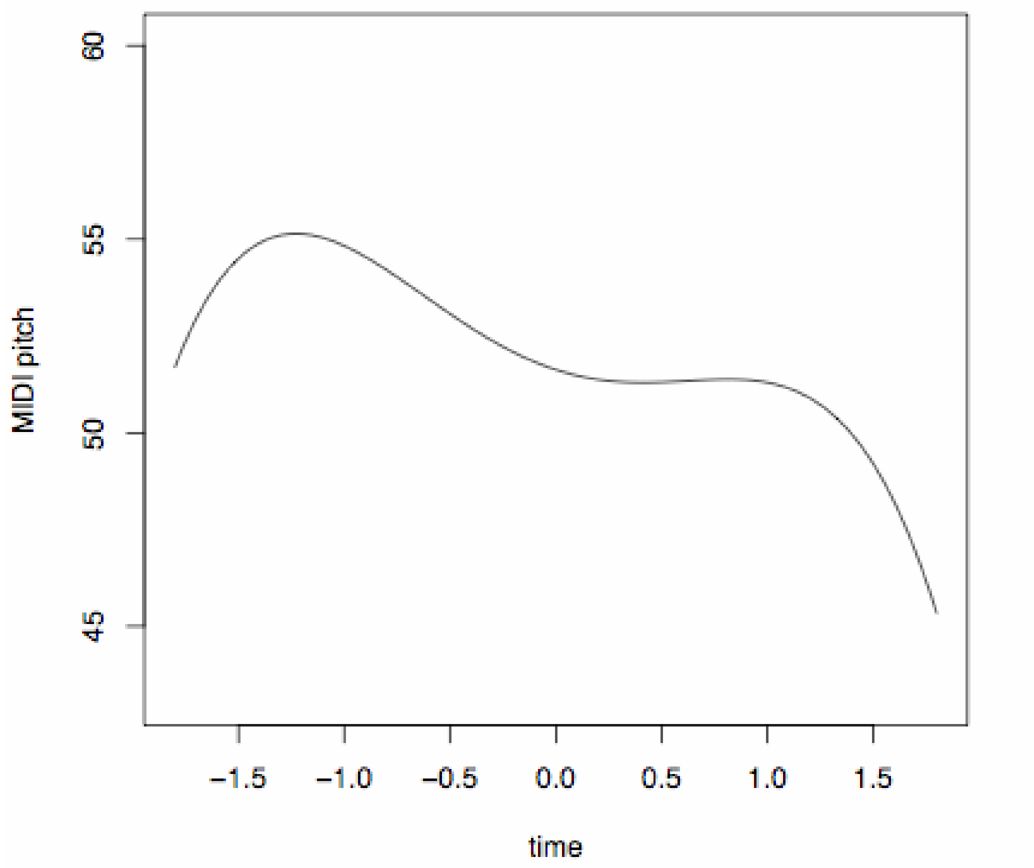

Fig. 15: Polynomial contour of *Ode to Joy*, 7th phrase

All the phrases are to a certain degree similar, in that their overall shape could be described as arch-like. However, they differ in length, the relation between the first and last note (overall rising vs. falling motion), and the range between the lowest and highest pitch in the phrase. We compare these phrases from Beethoven's *Ode to Joy* also to the first and second phrase of *Ah vousdirai-je, Maman* (Fig. 1), and the first phrase of the US national anthem (Fig. 8).

To measure the pairwise similarity between these phrases, we propose a new similarity measure based on Tversky's ratio model of similarity (1977) and a graphical interpretation of the polynomial curves. We assume that the area under the polynomial curve characterises the contour of a melodic phrase and that the overlap between the areas under the curves of two polynomials reflects their degree of similarity.

The polynomial functions $p(x)$ and $q(x)$ derived from melodies m and n as well as the time limits of the melodies to the left and the right of the origin $x_{p,-T}$, $x_{p,T}$, and $x_{q,-T}$, $x_{q,T}$ are the input to the procedure. First, the additive constant a in $p(x)$ and $q(x)$ is replaced by a value such that the lowest value of the polynomial curve within the limits is 0.[7] In the following we only give the equations concerning the transformations of $p(x)$. $q(x)$ is transformed correspondingly.

(5) $$a' = a - \min(p(x)); x \in \left[-T, T\right]$$

[7] This is only one possible way to transform two phrases into the same pitch range. Other possibilities which need to be explored in the future include offsetting the two phrases by their respective average pitches or by their central or key pitches.

Then we compute the areas below each polynomial curve by integrating the polynomial function in its limits.

(6) $$P = \int_{-T}^{T} p'(x)$$

Then we compute the two difference functions between the two polynomials:

(7) $$d(x) = p'(x) - q'(x)$$

This is done simply by subtracting each coefficient value of one function from its correspondent value in the other function. We denote the ranges of values for which $d(x)>0$ with R_i, each of which comprises a start and end point j and k.

We then integrate over all positive ranges and sum the resulting areas.

(8) $$D = \sum_{i=1}^{R} \int_{j}^{k} d(x)$$

This leaves us with four area values, *P, Q, D*, and *E*. The intersection between the areas under the two curves is then given by

(9) $$P \cap Q = P - D = Q - E$$

For computing a similarity between the two melodic phrases we make use of the ratio model of similarity originally proposed by Tversky (1977, p. 333) for measuring the similarity between two objects *a* and *b* via their feature sets *A* and *B*.

(10) $$s(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)}, \alpha, \beta \geq 0$$

Here $f(A \cap B)$ is a function that measures the salience or prominence of the features present in both melodies to the notion of similarity. By analogy, $f(A - B)$ and $f(B - A)$ measures the salience of the features only present in one of *a* and *b* respectively. The choice of the weights α and β sets the focus of the similarity comparison. A straightforward implementation of the ratio model in terms of our melodic contour representation is achieved by using the areas as values of the salience functions and setting $\alpha = \beta = 1$:

(11) $$s(m,n) = \frac{P \cap Q}{P \cap Q + D + E}$$

The similarity measure has a range of possible values between 0 (no similarity) and 1 (maximal similarity/identity).

Table 1: The similarity values obtained by the similarity measure in eqn. 11 for all pairs of melodic phrases of our test set.

	Ode_p1	Ode_p2	Ode_p5	Ode_p6	Ode_p7	US_nat_p1	Ah_p1
Ode_p1	1						
Ode_p2	0.51	1					
Ode_p5	0.56	0.67	1				
Ode_p6	0.53	0.68	0.95	1			
Ode_p7	0.73	0.37	0.48	0.46	1		
US_nat_p1	0.38	0.42	0.3	0.3	0.38	1	
Ah_p1	0.57	0.38	0.41	0.39	0.2	0.41	1
Ah_p2	0.6	0.47	0.74	0.74	0.49	0.31	0.33

The highest similarity value of 0.95 is obtained for phrase 5 and 6 from *Ode to Joy*, the full results being given in Table 1. Both phrases are one measure long, have a clear arch-like shape, the same pitches at start and end and a narrow interval range of 4 and 5 semitones respectively. The high contour similarity is reflected by the closeness of the polynomial curves and the very small area between them as depicted in Fig. 16.

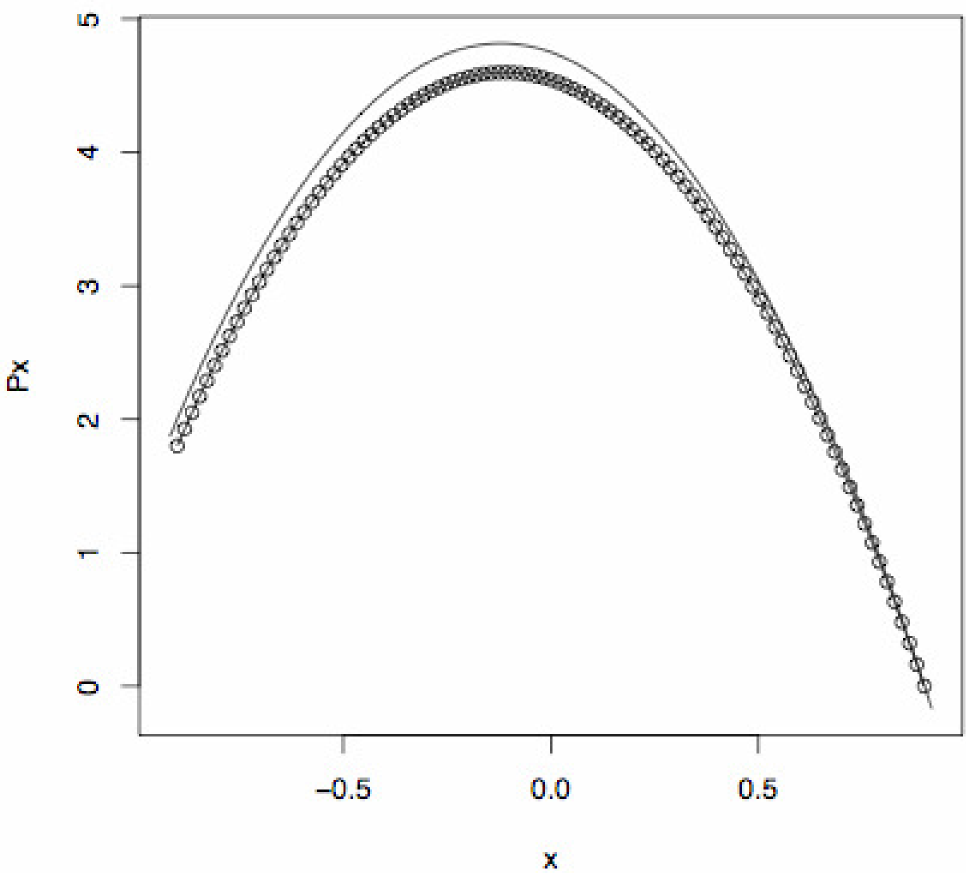

Fig. 16: Contour curves for phrases 5 and 6 from *Ode to Joy*.

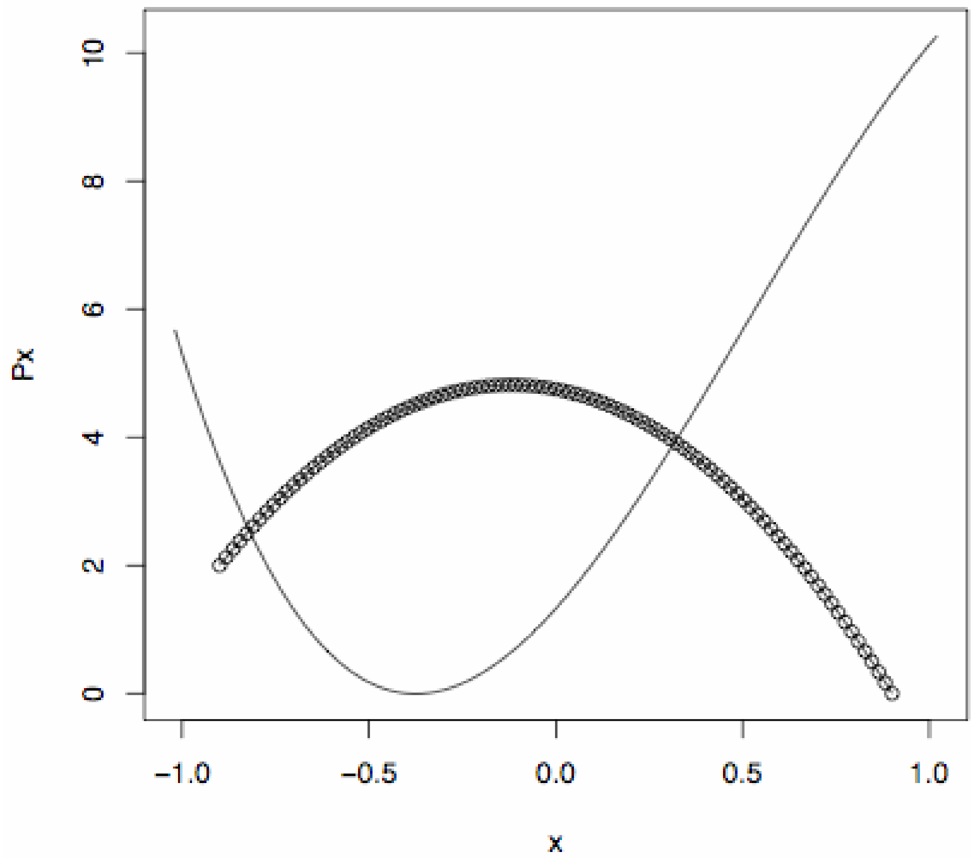

Fig. 17: Contour curves for phrase 5 from *Ode to Joy* and the first phrase of the US national anthem.

The other similarity values in table 1 between the 5 phrases from *Ode to Joy* cover a middle range between 0.37 and 0.73. Their common feature is the arch-like overall motion. Generally, higher values are obtained for phrases of approximately the same length (*s(p2,p5)*, *s(p2,p6)*) and phrases of approximately the same length and the same interval between start and end note (*s(p1,p7)*).

In contrast, clearly lower similarity values are obtained for the comparisons of the arch-like contours from *Ode to Joy* to the first phrase of the US national anthem, which describes something of a *J*-shape. Fig. 17 shows the small overlap between the areas under the curves of the latter phrase and phrase 5 from *Ode to Joy*.

5 Summary and options for future work

In this paper, we have introduced a new method for representing the contour of melodic phrases based on fitting a polynomial to the pitches of the phrase. In our approach, the melodic contour of a phrase is represented by a finite set of polynomial coefficients and the time limits that mark the start and end of the polynomial curve. We have given a detailed description of how the fitting is carried out in order to enable other researchers to implement their own version of this idea. We have explained several principled implementation decisions during the fitting process, where exploring alternatives might be a worthwhile enterprise. Among the decisions to be challenged are the length and duration limits of the phrases to be modelled by a polynomial, the decision not to normalise to unity with regard to pitch and time, and the z-transformation of the resulting coefficients.

We have presented two applications that make use of the polynomial contour representation and sketched the options for clustering large collections of melodic phrases. The first application allows one to estimate the probability density of a melodic contour with respect to a Gaussian mixture model created from a sample of 30,000 phrases randomly drawn from a pop song collection. The density model of phrase contours can be a very useful tool for a number of music information retrieval applications that deal with melodic data. In addition, cognitive models of music processing can be informed by density estimates of phrase contour assuming that the model density is closely related to the prevalence of a melodic contour (in popular music).

As a second application, we have defined a similarity measure that takes two polynomial curves as drawn within their corresponding limits and compares the areas enclosed between the curves. The degree of overlap between areas is then interpreted as the degree of (contour) similarity between the two melodic phrases.

Both applications have been tested by example in this paper and the results are plausible, easy to interpret, and intuitively convincing. However, a rigorous empirical verification of both applications, with large suitable datasets and more specifically selected qualitative examples, is necessary to evaluate the utility of the polynomial contour representation for theoretical, analytical, and empirical research on melodies. This is our next step.

Acknowledgements

We would like to thank Jamie Forth for comments and feedback on the manuscript of this article. Daniel Müllensiefen’s contribution to this work was supported by EPSRC research grant EP/D038855/1.

References

Baddeley, A. (1997). *Human Memory.Theory and Practice.* Hove: Psychology Press.

Boltz, M. & Jones, M.R. (1986). Does rule Recursion Make Melodies Easier to reproduce? If Not, What Does? *Cognitive Psychology*, 389–431.

Bradford, C. (2005). Heart & Soul: Revealing the Craft of Songwriting. London: Sanctuary.

Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its applications in the study of expressive timing. *Proceedings of the International Computer Music Conference* (pp. 17–22).

Cuddy, L.L. & Lyons, H.I. (1981). Musical Pattern Recognition: A Comparison of Listening to and Studying Structures and Tonal Ambiguities. *Psychomusicology, 1*(1), 15–33.

de la Motte, Diether. (1993). *Melodie: Ein Lese- und Arbeitsbuch.* München, Kassel: dtv / Bärenreiter.

de Nooijer, J., Wiering, F., Volk, A. &Tabachneck-Schijf, Hermi J.M. (2008). Cognition-based segmentation for music information retrieval systems. In C. Tsougras & R. Parncutt (Ed.), *Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM2008).*

Dowling, W.J. & Bartlett, J.C. (1981).The importance of interval information in long-term memory for melodies. *Psychomusicology, 1*, 30–49.

Dowling, W.J. & Fujitani, D.S. (1971).Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America, 49*(2, Part 2), 524–531.

Dowling, W.J. & Harwood, D.L. (1986). Melody: Attention and Memory. In Dowling, W.J. & Harwood, D.L. (Ed.), *Music Cognition* (pp. 124–152). Orlando: Academic Press.

Dowling, W.J., Kwak, S. & Andrews, M.W. (1995). The timne course of recognition of novel melodies. *Perception and Psychophysics, 57*(2), 136–149.

Edworthy, J. (1985). Interval and Contour in Melody Processing. *Music Perception, 2*(3), 375–388.

Eerola, T. & Toiviainen, P. (2004). MIR in Matlab: The MIDI Toolbox. Proceedings of the 5th International Conference on Music Information Retrieval.

Eiting, M. H. (1984). Perceptual Similarities between Musical Motifs. *Music Perception, 2*(1), 78–94.

Friedmann, M. L. (1985). A methodology for the discussion of contour: its application to Schoenberg's music. *Journal of Music Theory, 31*, 268–274.

Frieler, K., Müllensiefen, D. & Riedemann, F. (in press). Statistical search for melodic prototypes. Staaatliches Institut für Musikforschung.

Galvin, J.J., Fu, Q.J. & Nogaki, G. (2007). Melodic Contour Identification by Cochlear Implant Listeners. *Ear & Hearing, 28*(3), 302–319

Grabner, H. (1959). *Allgemeine Musiklehre.* Kassel: Bärenreiter.

Hindemith, P. (1940). *Unterweisung im Tonsatz: Bd. 1: Theoretischer Teil*. Mainz: Schott.

Huron, D. (1996). The Melodic Arch in Western Folksongs. *Computing in Musicology*, 10, 3-23.

Idson, W.L. & Massaro, D. W. (1976). Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition. *Perception & Psychophysics, 19*(2), 155–175

Idson, W.L. &Massaro, D.W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics, 24*(6), 551–565.

Jeppesen, K. (1935). *Kontrapunkt.* Leipzig.

Juhász, Z. (2009). Automatic Segmentation and Comparative Study of Motives in Eleven Folk Song Collections using Self-Organizing Maps and Multidimensional Mapping. *Journal of New Music Research, 38*(1), 71-85.

Juhász, Z. (2000). A Model of Variation in the Music of a Hungarian Ethnic Group. *Journal of New Music Research, 29*(2), 159–172.

Kachulis, J. (2003). *The Songwriter's Workshop: Melody.* Boston, MA: Berklee Press.

Kim, Y.E., Chai, W., Garcia, R. &Vercoe, B. (2000). Analysis of a Contour-based Representation for Melody.*Proceedings of the 1st Conference on Music Information Retrieval.*

Kühn, C. (1987). *Formenlehre der Musik.* München, Kassel: dtv / Bärenreiter.

Marvin, E.W. & Laprade, P. (1987).relating musical contours: extensions of a theory for contour. *Journal of Music Theory, 31*, 225–267.

McLachlan, G. & Peel, D. (2000).*Finite Mixture Models:* Wiley-Interscience.

Meyer, L. B. (1956). *Emotion and Meaning in Music:* University of Chicago Press.

Müllensiefen, D. & Frieler, K. (2004). Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgments. *Computing in Musicology, 13*, 147–176.

Parsons, D. (1979). *Directory of tunes and musical themes.* New York: Spencer Brown.

Pauws, S. (2002). Cuby hum: A fully Operational Query by Humming System. *Proceedings of the 3rd International Conference on Music Information Retrieval* (pp. 187–196).

Pearce, M. T., Müllensiefen, D. T, & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 9*, 1367–1391.

Perricone, J. (2000). *Melody in Songwriting: Tools and Techniques for Writing Hit Songs*. Boston, MA: Berklee Press.

Piston, W. (1950). *Counterpoint* .London: Victor Gollancz.

Polansky, L. (1996). Morphological metrics. *Journal of New Music Research, 25*, 289–368.

Quinn, I. (1999). The combinatorial model of pitch contour. *Music Perception, 16*, 439–456.

Rosen, C. (1976). *The classical style: Haydn, Mozart, Beethoven.* London: Faber.

Schaffrath, H. (1995). *The Essen Folksong Collection in the Kern Format.* Huron, D. (Ed.): Center for Computer Assisted research in the Humanities.

Schmuckler, M. A. (1999). Testing Models of Melodic Contour Similarity. *Music Perception, 16*(3), 109–150.

Shmulevich, I. (2004). A Note on the Pitch Contour Similarity Index. *Journal of New Music Research, 33*(1), 17–18.

Steinbeck, W. (1982). *Struktur und Ähnlichkeit: Methoden automatisierter Melodieanalyse.* Kassel: Bärenreiter.

Taylor, J. A. & Pembrook, R. G. (1984). Strategies in Memory for Short Melodies: An Extension of Otto Ortmann's 1933 Study. *Psychomusicology, 3*(1), 16–35.

Temperley, D. (2001). *The Cognition of Basic Musical Structures.* Cambridge, MA: MIT Press.

Thom, B., Spevak, C. & Höthker, K. (2002). Melodic segmentation: Evaluating the performance of algorithms and musical experts. *Proceedings of the 2002 International Computer Music Conference*.

Toch, E. (1923). *Melodielehre: EinBeitrag zur Musiktheorie.* Berlin.

Trehub, S. E., Bull, D. & Thorpe, L. A. (1984). Infants' perception of melodies: the role of melodic contour. *Child Development, 55*, 821–830.

Tversky, A. (1977). Features of Similarity. *Psychological Review, 84*(4), 327–352.

Zhou, Y. & Kankanhalli, M. S. (2003). Melody alignment and Similarity Metric for Content-Based Music Retrieval. *Proceedingsof SPIE-IS&T Electronic Imaging*. SPIE Vol. 5021, 2003, 112-121.