Musicae Scientiae
Discussion Forum 4A, 2007, 211-233

Melodic and contextual similarity of folk song phrases

TUOMAS EEROLA AND MICAH BREGMAN
Department of Music
University of Jyväskylä, Finland

• **ABSTRACT**

Various models of melodic similarity have been proposed and assessed in perceptual experiments. Contour and pitch content variables haven been favoured although music-theoretical and statistical variables have also been claimed to explain similarity ratings. A Re-analysis of earlier work by Rosner & Meyer (1986) suggests that simple contextual features can also be highly explanatory with more complex stimuli. A new experiment containing short melodic phrases investigated the effectiveness of several global and comparative variables. A multi-dimensional scaling solution indicated that both melodic direction and pitch range are highly relevant for making such similarity judgments and that the most salient aspects of melody when making similarity judgments are relatively simple context-dependent features.

Key words: music, similarity, melody, contour
Word count: 5840 words

In psychological research, experiments involving similarity judgments have been essential to uncovering many of the salient features involved in perception. Such a research program can also be an effective access point for understanding music perception — specifically the perception of melody.

Research in melodic similarity has been carried out by researchers working in music psychology and cognition, music information retrieval and music theory. Previous work in melodic similarity has been summarized by Deliège (2001a) and Hewlett & Selfridge-Field (1998). Deliège's article served as the introduction to a special issue of *Music Perception* devoted to melodic similarity. Many theorists and psychologists have attempted to abstract the most salient features of melody by analyzing music through extensions of Gestalt-based reasoning. These authors have applied theories initially developed to explain visual perception to the domain of

sound and music. Researchers in music cognition have attempted to use empirical evidence to link similarity ratings given by experimental participants with features of the compositions used as stimuli. Engineers, computer scientists, and digital librarians have attempted to build workable similarity ranking systems whose primary focuses are usability and efficiency rather than the explanatory power of the underlying model.

Among the most frequent applications for similarity measures is the retrieval of music from large databases based on content queries. An intuitive human interface for a music information retrieval (MIR) system would allow the user to "hum" in an approximate version of the requested composition. The database would then retrieve a ranked list of results ordered by perceptual similarity to the input melody. A significant amount of recent work has concentrated on developing such systems (Downie, 2003; Wust & Celma, 2004).

Music theorists have largely focused upon developing techniques to understand grouping behaviour as it relates to melodic perception. One such grouping approach was the proposition of archetypal melodic gestures extracted from music by common-practice era composers (Rosner & Meyer, 1982, 1986). It has recently been shown, however, that the perceptual effects of some melodic archetypes are more effectively explained by much simpler statistical parameters (von Hippel, 2000; von Hippel & Huron, 2000). Implication-realization models provide an additional framework within which grouping algorithms can be developed (Lerdahl & Jackendoff, 1983; Narmour, 1992). Several musicologists have proposed that cue abstraction, which includes the categorization and organization of musical motives, may provide important information for similarity perception (Cambouropoulos & Widmer, 2000; Cambouropulos, 2001; Deliège, 1996; Deliège, 2001b; Koniari, Predazzer, & Melen, 2001; Zbikowski, 1999).

Several different approaches to building quantitative models of melodic similarity have been proposed, most of which have relied first on abstracting melodic information and then on somehow measuring the similarity between these abstractions. The simplest methods can be described as string-matching, where a melodic query string is input and melodies which contain this string or its transformations are located (Downie, 2003; Lemström & Ukkonen, 2000). Aggregate statistical data has also been used to model listener's similarity judgments (Eerola, Järvinen, Louhivuori, & Toiviainen, 2001). Other authors have tested the cognitive viability of a range of different similarity models, indicating that several different aspects of melody may prove applicable to modelling similarity judgments (Müllensiefen & Frieler, 2004; Toiviainen & Eerola, 2002).

The most wide-spread and intuitive abstraction of melody is contour. Contour abstractions seek to limit the specificity of melodic information while retaining the essential shape of a melody thought by Schoenberg (1975) and others to have perceptual relevance. Contour representations vary in the type of information they limit and in the extent to which it is reduced. Some methods simply discard

rhythmic information, while others go as far as discarding rhythmic information and reducing all intervallic information to either ascending, descending or unchanging (Uitdenbogerd & Zobel, 1999).

Melodic contour has been used as a representation by researchers studying many aspects of melodic perception for more than thirty years (Dowling, 1971, 1994; Idson & Massaro, 1978; Massaro, Kallman, & Kelly, 1980; Monahan & Carterette, 1985; White, 1960). This body of work has shown it to be a particularly robust representation of melody. Work has also been done specifically on the development of an ideal representation and distance measurement method for contour (Marvin & Laprade, 1987; Polansky, 1996; Quinn, 1997, 1999). Contour has been utilized in similarity measures as a representation of both tonal and atonal melodies, including serial tone rows (Schmuckler, 1999).

More recently, neuroscientists have begun to use event-related potential (ERP) experiments to understand the role that contour may play in melodic perception and to distinguish this from the role played by intervallic information (Brust, 2003; Tervaniemi, Rytkönen, Schroger, Ilmoniemi, & Näätänen, 2001; Trainor, Desjardins, & Rockel, 1999). This work has contributed evidence that melodic contour is more quickly accessible than intervallic information and contour changes are detectable even when attention is not devoted to the music listening task. They have also given additional weight to earlier psychological results indicating that alterations to the contour of a melody are more destructive to its recognition than alterations to its pitch content that do not affect contour.

One of the difficulties of developing a melodic similarity model is effectively distinguishing the boundaries between phrases. Comparing two melodies phrase-wise allows contour analysis in particular to be applied effectively. While certain information, such as pitch class distribution and meter may be useful when considered in aggregate, breaking melodies down into "chunks" is a natural process of both music listening and music production so its inclusion as a necessity in a perceptually-consistent model of melodic similarity is clear.

Although numerous attempts have been made to develop reliable automatic segmentation systems, it remains a difficult problem (Ahlbäck, 1997; Bod, 2002; Bregman, 1999; Cambouropoulos & Widmer, 2000; Tenney & Polansky, 1980). Rather than focus on testing melodic similarity and melodic segmentation models simultaneously, in this experiment we chose to focus on similarity judgments between pre-segmented melodic phrases.

THE EFFECT OF CONTEXT ON SIMILARITY FORMATION

In the common *fixed-set approach*, similarity of two objects is evaluated from a set of representational features by computing certain types of distances between them. In addition, the features might be weighted in terms of their relative importance. This

fixed-set approach is commonly assumed and even required for a stable similarity formation (Tversky, 1977). However, context has a huge influence on feature selection and weighting decisions (Goldstone, Medin & Halberstadt, 1997; Rips, 1989). Tversky himself (1977) provided evidence for an *extension effect*, in which features influence similarity more when they vary within an entire set of stimuli. In other words, different frames of comparison are created depending on the items that are present in the rated set (Medin *et al.*, 1993). To summarise, according to Goldstone *et al.*, (1997, p. 238), "similarity is not simply a relation between two objects; rather, it is a relation between two objects and a context."

Although studies in music perception usually adopt a fixed-set approach to similarity, contextual effects have often been observed. Deliège's (2001a) theoretical notion of *imprint formation*, and subsequent empirical observations, demonstrate that the abstracted cues are dependent on the musical material. Lamont and Dibben (2001) provide evidence that musical context shapes the choice of features used in similarity formation, which in their study, is evident in differences in the choice of surface features for the similarity of Schoenberg and Beethoven excerpts. Similarly, McAdams *et al.* (2004) found that listeners capitalized on surface similarities when rating similarities between segments of a contemporary piece. Moreover, the variety of different musical features used in assessing similarity in the previous music perception literature probably reflects the effects of contexts in musical similarity (*e.g.*, Serafine, Glassman & Overbeeke, 1989; Monahan & Carterette, 1985). For example, Rosner & Meyer (1982, 1986) investigated the role of underlying deep or reductional structures in categorization and similarity which have been emphasized in music theory. To demonstrate how this essentially fixed-set approach actually reveals large contextual effects, we briefly investigated the empirical MDS similarity data provided by Rosner & Meyer (1986). Using MIDI representations of the stimulus melodies rather than the instrumental recordings used in the original experiment, we found that their coordinates in the MDS solution correlated highly with very simple melodic features. In their first experiment, Rosner and Meyer compared melodies that were chosen as prototypical "linear" melodies with those chosen as prototypical of "changing-note" melodies. These archetypal classifications are based upon a very high-level analysis of melodic structure.

According to Rosner and Meyer, changing-note melodies are defined by a progression from the tonic to the leading tone followed by the second scale degree and then the tonic again. Melodies classified as linear are those whose structure can be reduced to ascending and descending seconds and thirds. Although Rosner and Meyer's cluster analysis of the MDS results do show some tendency by participants to separate the melodies into groups, the results could also be subject to a much simpler analysis that doesn't rely upon a theory of melodic archetypes. As illustrated in Figure 1, the y-coordinate of the MDS solution can be correlated to the standard deviation of the melodies pitch height (r = .946, n = 12, p < .001) and the x-coordinate to the interval between the first and last pitch of the melody (r = .858,

n = 12, p < .001). These very high correlations suggest that simple melodic features may play an important role in melodic similarity judgments, because these features provided the context in which the melodies varied systematically (*i.e.*, Tversky's extension effect). In other words, the listeners paid attention to the most salient variation within the set of stimuli, which in this case was variation in pitch height and pitch direction. The results from a study by Lamont and Dibben (2001) — although they focused exclusively on such surface features as dynamics, articulation and texture — also show similar tendencies in their correlation of MDS solution axes with variables such as melodic direction and other surface features of Beethoven and Schoenberg excerpts. With this in mind, we attempt to investigate melodic similarity using a more constrained yet natural set of stimuli.

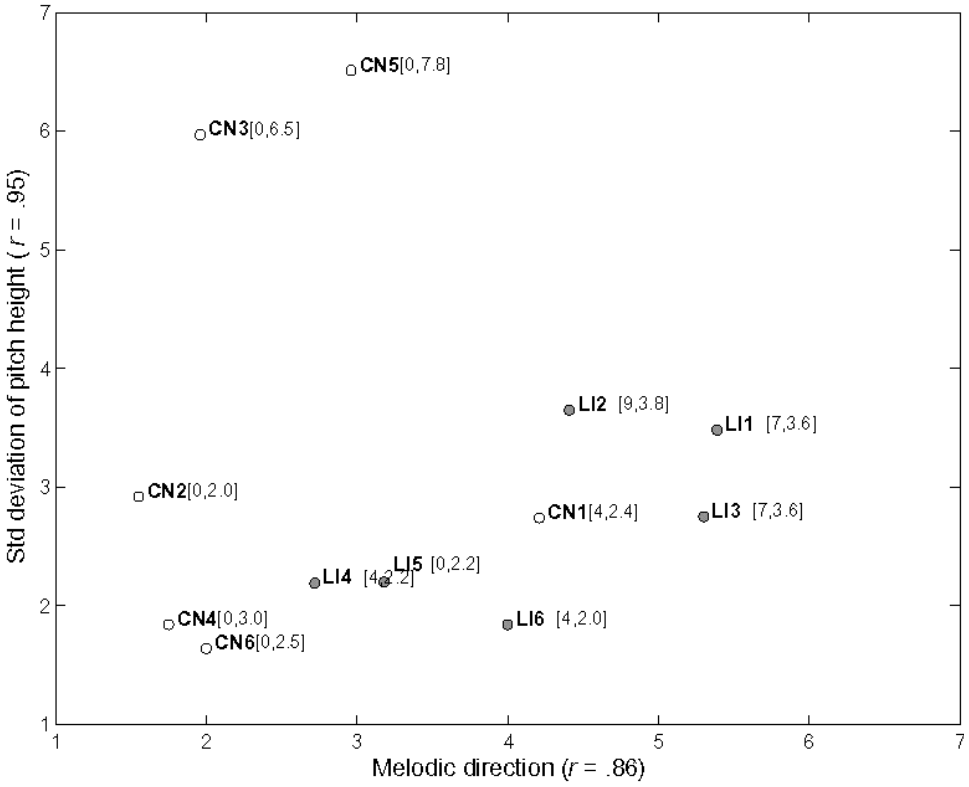

Figure 1.

Replication of Rosner and Meyer (1986) data. The point labels indicate the melody name given in their article while the numbers in brackets indicate the [x, y] values for melodic direction and standard deviation of pitch height.

RATIONALE AND HYPOTHESIS

In summary, a number of researchers have developed models of melodic similarity for widely disparate purposes. Some have relatively little perceptual relevance while others have been distant in their applicability to real musical stimuli. Some of the other previous experiments have relied on complex theories of archetypal melody groups or primary contour "building blocks" (Adams, 1976). Although attractive simplifications of complex grouping principles, it can often be difficult to successfully link empirical data with such approaches.

Previous experiments have used both real musical stimuli and artificially generated melodies. Although using real musical stimuli is tempting, the segmentation difficulties make it particularly challenging to separate similarity effects from segmentation effects. Rather than generating artificial phrases for this experiment, however, we used a cross-section of the Essen folk melody collection (Schaffrath, 1995) so that the results could be more clearly applied to musical situations. By developing a simpler approach to melodic similarity perception based upon phrase similarity rather than entire melodies, we expected to access the simpler contextual predictors of similarity more easily. Simple similarity predictors might allow us to better understand the salient factors and context-dependency that contribute towards melodic perception.

SIMILARITY PREDICTORS

The theoretical models discussed in the introduction suggest the existence of several types of predictor variables which are able to model listener's pair-wise phrase similarity ratings. The predictor variables we chose exhibit a range of different types of melodic representations and therefore should provide several different access points for understanding similarity ratings. We employ two kinds of similarity predictors: pair-wise and global similarity predictors. In the first category of predictors, each variable is named by a single property of a melody, although it is calculated as the *difference* in this characteristic between a pair of melodies. Larger values of these variables therefore correspond to a higher degree of distance, or dissimilarity between the melody pair. The second category, global similarity predictors, consists of simple, one-dimensional features of a single phrase, not of a phrase pair. Their purpose is to uncover possible similarity rating strategies emerging from a multidimensional scaling solution of the similarity ratings. As these are summary measures of pair-wise predictors they will be described in conjunction with each pair-wise predictor. All variables were created in Matlab using MIDI Toolbox (Eerola & Toiviainen, 2004).

• ***Contour*** (pair-wise predictor) is a measure of the geometric distance between a melody pair's melodic contour vectors. Contour vectors were created by taking 10

equidistant samples of pitch height for each melody (see Figure 2 for an example) and using the euclidean distance function to calculate the distance between them. Previous work by Ó Maidin (1998) suggested that geometric distance could be a useful way of measuring differences between contour vectors. There are two global measures of contour: *mean pitch* and *melodic direction*. *Mean pitch* indicates simply the mean pitch of a melody, coded in terms of MIDI pitch height (middle c = 60). *Melodic direction* is the interval (in semitones) between the first and last note of the phrase.

• ***Pitch content*** (pair-wise predictor) gives a simple indication of the differences in pitch content between two melodies. For each melody, the proportion of each pitch-class weighted by Parncutt's (1994) durational accent model was represented in a distribution. The taxi-cab distance (sum of the absolute value of the difference between two vector's components) was measured between the two distributions. Two melodies with identical distribution of pitch content would have zero pitch content difference, whereas two melodies with no pitches in common would have a value of two. Four global predictors of pitch content are proposed. *Entropy of pitch content* describes the orderliness of the pitch distribution by calculating the entropy of the distribution according to the following equation

$$H = \frac{-\sum_{i=1}^{N} p_i \log_2 p_i}{\log N}$$

where H is the relative amount of information conveyed, N is the number of possible unique states (in this case pitches) in the repertoire and p_i is the probability of occurrence of state i in the repertoire. This measure has been previously applied to various musical contents (Knopoff & Hutchinson, 1983; Snyder, 1990). *Range* indicates the total number of semitones between the lowest and highest pitched notes of each phrase. In order to deal with the pitch content in a slightly more specific manner, the role of implied harmonic chord tones was covered with two predictors: *tonic pitches* and *dominant pitches*. They are calculated as the durational proportion of tonic or dominant pitches in the phrase, which clarifies the implied harmonic content.

• ***Interval content*** (pair-wise predictor) is similar in concept to pitch content, although it was measured as the taxi-cab distance between two interval distribution vectors. Thus two melodies which were entirely stepwise ascending would have a very low value of interval content distance whereas two melodies with differing interval content would have a higher value. Three global predictors of interval content are proposed: *mean interval size*, *stepwise movement* and *triadic movement*. The first of these needs no further definition. *Stepwise movement* is a measure of the

tendency of the phrase to contain stepwise motion. It is coded as the proportion of intervals in the phrase which are either major or minor seconds and *triadic movement* is a measure of the tendency of the phrase to contain movement by thirds. It is coded as the proportion of intervals in the phrase which are either major or minor thirds.

• ***Contour periodicity*** (pair-wise predictor) is based on a fourier spectral analysis of the contour vector. Fourier analysis is a mathematical method which allows signals to be represented as the sum of a series of sinusoidal functions of varying frequencies and phases. By selecting only particular components of the discrete fourier transformation, the amplitude and the phase of the cyclic functions at particular frequencies can be determined. These amplitudes can be interpreted as the degree of periodicity of the melodic contour at particular frequencies. Thus if a melodic contour has a strong cyclical structure with two cycles over the course of the melody, the fourier analysis will indicate a relatively strong amplitude at the corresponding frequency (see Figure 2 for an example).

Previous work by Schmuckler (1999) applied both the phase and amplitude information present within fourier analyses to melodic contours. Schmuckler develops the theory behind application of fourier analysis and shows that spectral amplitude information was among the most effective predictors of similarity when derived from melodic complexity ratings. Its application to pair-wise similarity ratings has, however, not been previously tested.

We calculated the contour periodicity variable using 10-component contour vectors constructed with 10 equidistant pitch height samples for each melody. After processing using the discrete fourier transform algorithm, the resulting 10-component complex-valued vector was reduced to four components, as the first component includes information related to the constant non-cyclic part of the fourier analysis and the last five components of the vector are simply a mirror image of the first. The absolute value of each of the 4 components of the complex vectors were taken and distances between these vectors compared using the taxi-cab distance metric. As a global predictor of oscillating contour structure, we use *mean contour periodicity*, which is the mean frequency obtained from the spectrum of contour periodicity. A low mean frequency indicates large period structure whereas a high value suggests a faster oscillating pattern within a phrase.

STIMULI SELECTION

To obtain a realistic, representative, yet varied sample of melodic phrases for the experiment, a large collection of folk melodies was organized into prototypical phrases by means of a self-organizing map (SOM). The SOM is an artificial neural network that simulates the process of self-organization in the central nervous system with an effective numerical algorithm (Kohonen, 1997). It consists of a two-dimensional planar array of simple processing units, each of which is associated with a reference vector. During the necessary learning session, the SOM performs

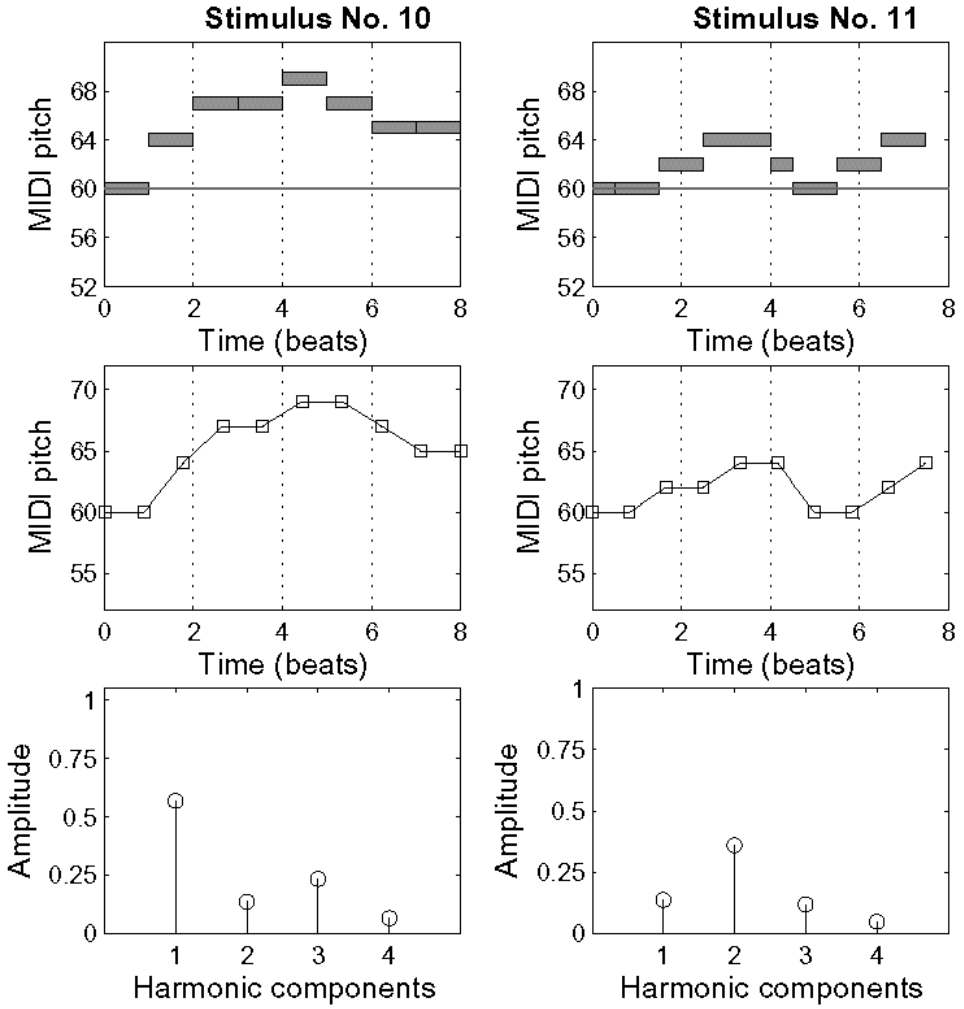

Figure 2.

Example of stimulus 10 and 11 using pianoroll representation (topmost panels), contour vector (middle panels) and fourier spectral analysis of the contour vector, i.e., contour periodicity representation (lower panels).

clustering, attempting to represent the training set by a low number of prototypes. As a result, it forms a non-linear projection that optimally approximates the distribution of the data where similar vectors are mapped close to each other.

To train the SOM, we used a collection of European folk melodies (n = 6236) from the Essen collection (Schaffrath, 1995), amounting to 35 808 explicitly marked phrases. The contours of the phrases were extracted by sampling the pitch height of the phrase at 32 equidistant sampling points using nearest neighbour interpolation. Also, the rests were omitted and where rests occurred the previous pitch was sampled.

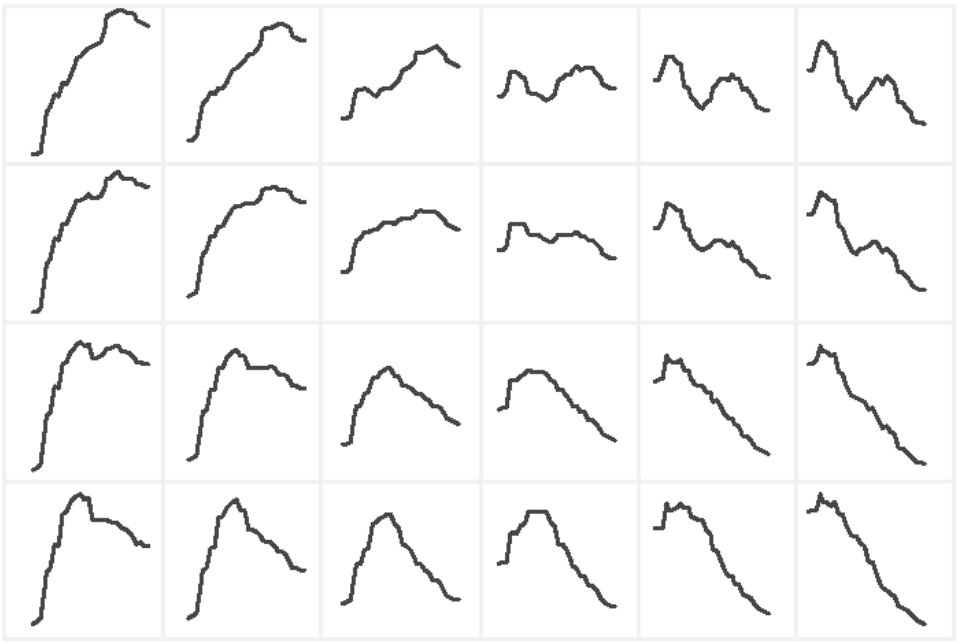

Figure 3.
4 × 6 grid of prototypical contours from the SOM trained with 35 808 phrases from the Essen collection (Schaffrath, 1995).

Next, this contour vector was normalized (mean pitch height = 0, standard deviation = 1). A SOM containing 4 × 6 grid was trained with this data, yielding 24 prototypical contour vectors (Figure 3). The size of the grid was limited, as in pair-wise rating experiments, the number of pairs increase very rapidly with the number of stimuli at a rate of $n\,(n-1)\,/\,2$.

The phrases for the experiment were chosen from the Essen collection to correspond to the prototypical phrases present in the collection and provided by the SOM. This correspondence was measured by ranking phrases in each cell by the quantization error between the individual phrase vector and the prototypical phrase vector. The highest-ranked phrase was chosen from each of the SOM cells that fulfilled several simple criteria. Only melodies in major mode and simple duple meter (2/4 and 4/4) were included. Further, the number of note events was restricted to between 6 and 15 and the lengths of the phrases were restricted to between two and four seconds. No melodic phrases which contained rests were included.

The melodies were transposed to reduce the potential for key-context effects. In each case the phrase was transposed so that minor seconds present in the melody were consistent with interpretation of the melody in the key of C-major. Only melodies whose interval structure allowed them to start on the tonic were included. In 4 of the SOM cells, none of the first 500 phrases in the ranked list qualified as

stimuli. This was because a tendency for minor mode phrases within these cells or phrases which were unable to begin on the tonic. Melody 8, however, was included as it provided an excellent example of a phrase with a very flat contour. Moreover, as it only included two pitches we did not expect it to elicit a strong sense of tonality. Similarly, some of the candidate phrases could be interpreted in another key or even meter due to isolation from other parts of the melody and absence of harmonic context. This resulted in 20 phrases (plotted in pianoroll notation in Figure 4) each of which consist of between 4 and 9 note events and is 2.4 to 4.4 seconds in duration with a median of 3 seconds[1].

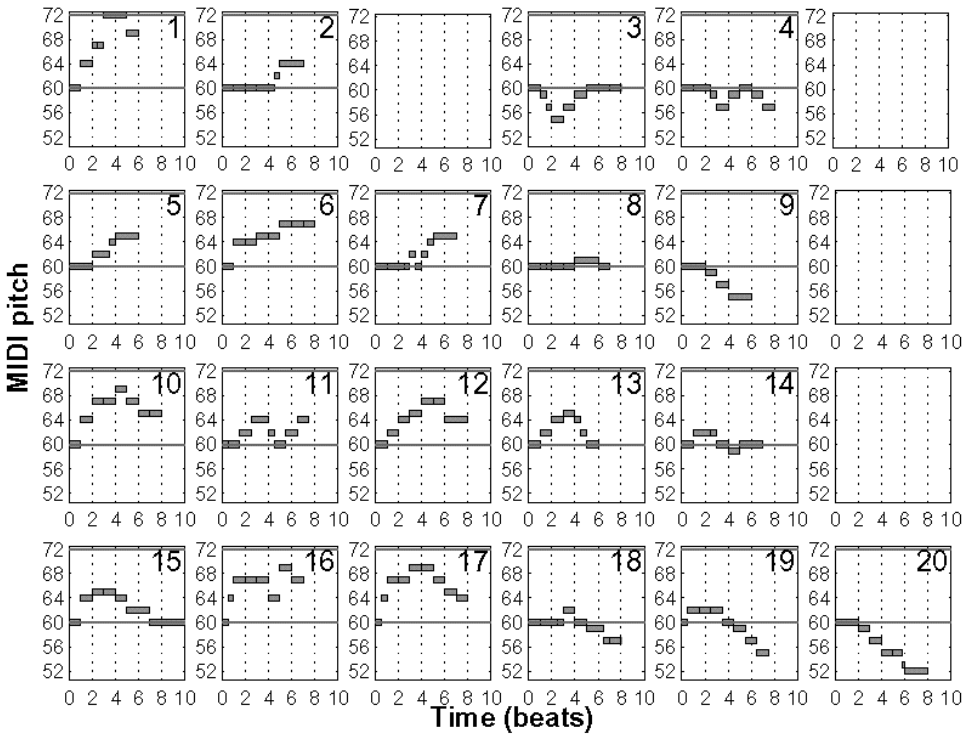

Figure 4.
20 stimulus phrases used in the experiment plotted using pianoroll representation.

(1) List of stimuli where the names refers to the Essen collection and number after the colon indicates the phrase number: (1) deut4568: 1, (2) deut2044: 1, (3) deut526: 5, (4) deut2556: 3, (5) deut230: 2, (6) deut1034: 1, (7) deut376: 2, (8) deut540: 1, (9) deut0699: 1, (10) deut4974: 3, (11) deut1507: 3, (12) deut1699: 2, (13) deut4736: 2, (14) deut152: 4, (15) deut4389: 3, (16) deut0899: 1, (17) deut1413: 2, (18) deut3623: 3, (19) deut1507: 2, (20) deut1758: 1

EXPERIMENT

STIMULI AND EQUIPMENT

All 20 phrases were presented at 150 quarter-note beats per minute and synthesized using a piano timbre via Cakewalk Home Studio 2002 and Virtual Sound Canvas running on an IBM-compatible computer. The interstimulus duration between paired phrases was 1400 ms which consisted of a 500 ms pause after the stimulus followed by a 400 ms series of two percussive clicks indicating the end of the first stimulus and the beginning of the second stimulus and an additional 500 ms pause. All stimulus pairs were presented using Millisecond Software's Inquisit 1.33 via a NAD 7120 Amplifier and pair of B&W DM100 speakers.

METHOD

As similarity rating is a time-consuming and exhausting task, we attempted to reduce the number of pair-wise ratings required. It has been proposed that overall MDS results are unaffected by some reduction in the number of comparisons made. Up to a 30 percent random omission of possible pairings has been shown to have little effect on results when subjected to MDS analysis (Tsogo, Masson, & Bardot, 2000).

Rather than eliminating stimuli pairs completely, we decided to omit pairs cyclically so that the rating task for individuals was reduced, but no pairs were omitted entirely. This meant that 25 percent of the phrase pairs were randomly discarded and 4 versions of the rating experiment were created, each containing 75 percent of the possible pair-wise comparisons with a unique 25 percent omission. In this manner, 4 participants provided 3 ratings of each melody pair.

PARTICIPANTS

Participants for the experiment were volunteers from the University of Jyväskylä music department. Of the 22 participants whose data was included in the analysis, 9 were men and 13 women. The mean age of the participants was 23.9 years with standard deviation of 4.6 years. They had music as a hobby for an average of 16.4 years (SD=4.2 years) and had studied formally for an average of 9.8 years (SD=4.5 years).

PROCEDURE

Prior to beginning the experiment, participants read instructions on the computer screen where they were told their task was to rate the similarity of pairs of melodies on a scale from 1-9 where 1 corresponded to "very dissimilar" and 9 to "very similar". The participants made the ratings using the numeric keys on the computer keyboard, with visual confirmation of their choice on the screen. After confirming their choice, the next stimulus pair was presented immediately. Participants were informed that all melodies were taken from the same collection of European folk songs and that all of them would be played using a piano timbre. They were encouraged to use the entire range of the rating scale.

Participants were first presented with 5 practice trials in order to familiarize themselves with the procedure and the types of melodies they would hear. These practice trials were followed by 4 experimental blocks in which data was collected. The melody pairs presented in the practice trials were chosen from the 25 percent of melodies excluded in the blocks. The first three blocks each consisted of 35 randomly ordered phrase pairs, while the final block contained the remaining phrase pairs. Participants were free to take breaks between blocks.

All participants were tested individually in a sound-isolated room. After the experiment, the participants completed a musical background questionnaire and were debriefed of the purpose of the experiment. The total duration of the experiment including the instructions, practice trials and questionnaire, was about 45 minutes.

RESULTS

Of the 24 participants, the data of 2 were discarded on the basis of low inter-subject correlations. Only data from participants whose inter-subject correlations corresponded to $p < .05$ were retained. Owing to a programming error, only 8 subjects rated pair-wise comparisons with stimulus melody 1. However, the standard deviation in these ratings (1.55) was not higher than the standard deviation in the rest of the comparisons (1.62). This implies that although fewer ratings were given comparing melody 1 to the other stimuli, the ratings given provided sufficient information for inclusion in our data[2].

Half of the participants were presented with melody pairs in one order, and the other half with the pairings reversed. The two groups were highly correlated ($r = .781$, $n = 190$, $p < .001$) and showed no significant difference in their means (t (416) = 1.54, ns) so their results were considered together in all analyses. For convenience of analysis, all ratings were reversed, so that higher ratings indicated a greater degree of dissimilarity between the melodies. Likewise, all variables were constructed so that larger values indicate a greater degree of dissimilarity (or distance) between the stimulus pairs.

The first analysis approach consisted of constructing variables which replicate the entire pair-wise comparison matrix. These techniques used direct pair-wise feature comparison of the melodies which were compared with the similarity rating matrices derived from the empirical data using correlation coefficients. In other words, each stimulus pair was assigned a single value corresponding to their difference under the

(2) The results remain identical if the melody 1 is removed from the analysis. In the first round of analysis, the regression equation explains 5 % more of the variance without melody 1 with the same predictors as with all the melodies. In the second round of analysis, both MDS solution and correlations remain virtually the same with or without melody 1 (mean change in the correlations is – 0.002).

Table 1
Correlation Matrix of the Pair-wise Similarity Predictors and Similarity Ratings (*N* = 190)

Variables	*Contour*	*Pitch content*	*Interval content*	*Contour periodicity*	*Range*
Ratings	.605**	.689**	.353**	.473**	.389**
Contour		.804**	.426**	.760**	.480**
Pitch content			.313**	.507**	.530**
Interval content				.549**	.316**
Contour periodicity					.251*

* $p < .05$, ** $p < .001$

given variable. Table 1 shows the correlations between listeners' similarity ratings and pair-wise similarity predictors.

Participant's empirical pair-wise similarity ratings correlated quite highly with several of the predictor variables indicating that even a single predictor can have a fair amount of explanatory power. The highest single correlation occurred between the difference in pitch content of the melodies and the listener's ratings (r = .689). This indicates that listeners were sensitive to the pitch content of the melodies when making similarity judgments. It must be noted, however, that pitch content and contour were highly co-linear (r = .804) indicating that in short phrases such as the stimuli used in the experiment, melodies differing in pitch content also tended to differ in contour. The high correlation between geometric contour distance and listener's ratings (r = .605) is not surprising. Contour is certainly the most widely applied abstraction of melody, and has been previously introduced as a successful predictor for similarity judgments. The high correlations between the variables could also be interpreted as reflecting the compositional practices of this folk song material, where multiple cues are used to maximize memorability of phrases. In order to check for primacy and recency effects, where the initial or the last part of the stimulus is better remembered than other parts, the beginning and the end of the contour representation were weighted separately using an exponential function prior to calculating pair-wise distances. A systematic variation of the exponent did not improve the correlation between this weighted contour variable and the ratings.

The large number of highly significant correlations between the variables suggests that although different melodic features were measured, there was a high degree of co-linearity between several features. Contour, in particular, correlated highly with a number of the other predictor variables, indicating that some of the other variables also served as a measure of contour difference. Correlation between contour and contour periodicity (r = .760) is expectedly high, as fourier transformation of contour is another representation of contour information.

Further analysis using standard multiple regression analysis was conducted to reveal how individual pair-wise variables relate to the similarity ratings. All variables explain up to 51 % of the variance in participants' similarity ratings (R = .72, F (4,185) = 48.35, p < 0.001), and three of the four predictors contributed statistically significantly to the regression equation. These were contour (β = .45, p < .001), pitch content (β = .40, p < .001) and interval content (β = .17, p < .01). Although there is co-linearity amongst the predictors (especially contour and pitch content, r = .804), three predictors still contribute individually to the listeners' similarity ratings, suggesting the underlying similarity is a combination of several features that were at least partially measured by the available predictors.

In order to further indicate the presence of simple contextual predictors in similarity perception, we also used a two-dimensional multidimensional scaling solution (MDS) to reduce the dimensionality of the data (Kruskal & Wish, 1978). Multidimensional scaling is an algorithm which allows an n-dimensional map to be constructed from pair-wise distance data. Each stimulus is assigned a position on this map such that the distances between it and the other stimuli correspond as closely as possible to the pair-wise ratings. Effectively, this not only dramatically reduces the dimensionality of the data, but may allow additional avenues for analysis. Some resolution is lost, however, as the data necessarily undergoes a certain amount of stress in the process such that not all data is preserved in the transformation. In this data, a two-dimensional solution yielded a Kruskal stress value of .19, meaning that 81 % of the original data was accounted for by the solution. The three-dimensional solution was only marginally better (Kruskal stress = .13) and hence the two-dimensional solution was used to visualize the data (Figure 5).

We then correlated the X and Y axis of the MDS solution with the global melodic features (Table 2). This analysis produced a very similar explanation of the data to that uncovered in the re-analysis of Rosner and Meyer's work. In particular, the x-coordinate of the MDS correlated extremely highly with melodic direction (– .866), mean pitch height (– .963) and mean interval size (– .807) whereas y-coordinate correlated with the proportion of tonic pitches (– .668), pitch content (.657) and range (.654). To put it simply, the x-axis corresponds to melodic direction while the y-axis corresponds to melodic range or pitch variation. It is worth noting that overall contour and pitch content provide significantly more explanatory power together than one or the other alone. This emphasizes the multi-dimensionality of contextual similarity predictors and provides further evidence that simple predictors can explain a substantial portion of the similarity ratings of melodies. This explanation does not imply that these features are the most vital for similarity formation in general but that they represented the most salient variation within this particular stimulus set. These particular predictors, nonetheless, are the ones that the reviewed literature usually assigns as the most robust dimensions of melodic similarity in Western tonal music, and hence some combination of these few predictors, is likely to be highly relevant in any given melody. Whether the same predictors would be relevant in

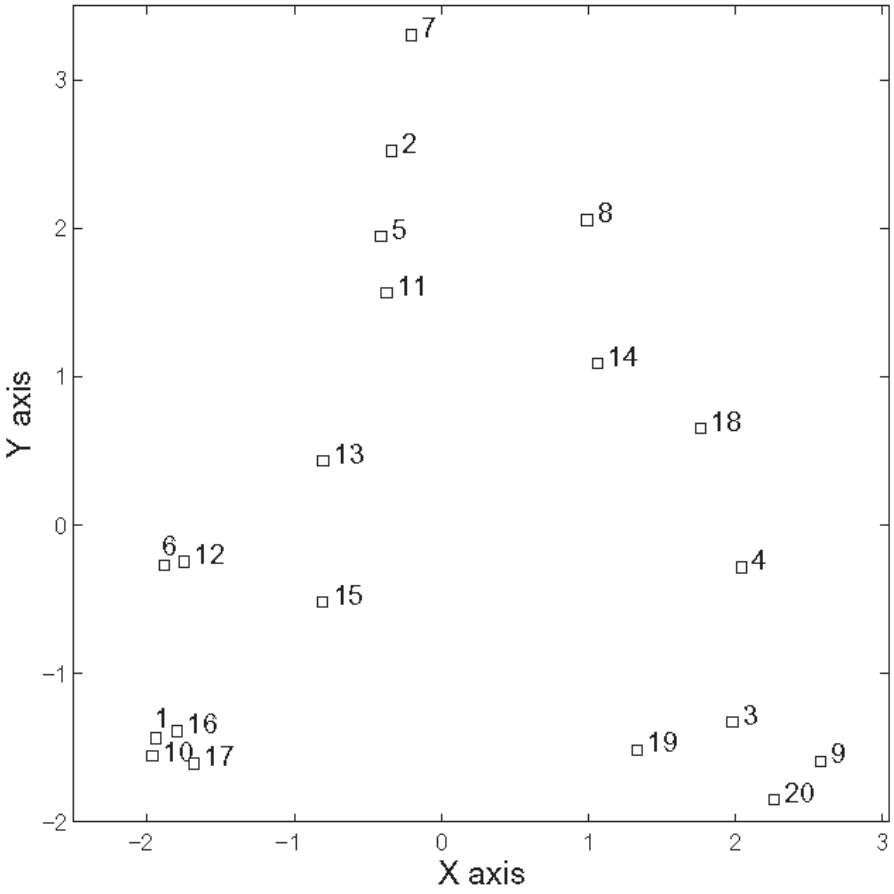

Figure 5.
Multidimensional scaling solution for the similarity of the 20 stimulus phrases.

musical similarity in non-Western traditions, is a question that warrants further research.

It is also possible that weighting the predictors across the melody in a certain manner would improve the fit between the predictors and the data. There are at least two ways of doing this: by increasing the weight of events that occur on beats of differing agogic stress and by increasing the weight of the phrase beginnings. The first principle is well-known in the literature (Dibben, 1994; Palmer & Krumhansl, 1990) and has been shown to increase the melodic similarity by Eerola and his colleagues (2001). The second principle is more particular to the pair-wise rating methodology, where the melodies are repeatedly heard and instantly compared. In this type of paradigm, listeners may tend to prioritize the early part of the phrase in their similarity decisions, although the importance of phrase beginnings is acknowledged in other contexts as well (Chiappe & Schmuckler, 1997; Schulkind,

Posner, & Rubin, 2003). An exponential weighting of the phrase beginnings and endings was performed to assess this but it did not enhance the fit between the contour predictors and the data. Nevertheless, a more thorough investigation of the contextual dependency and similarity features is necessary in order to assess how the two components build upon each other.

Table 2
Correlations Between MDS Solution Axes and Global Predictors (*N* = 20)

Global predictor	X coordinate	Y coordinate
Mean pitch	-.963**	.077
Melodic direction	-.866**	-.228
Entropy of pitch content	-.124	.657**
Range	-.540*	.654**
Tonic pitches	.549*	-.668**
Dominant pitches	-.705**	.434
Mean interval size	-.807**	-.304
Stepwise movement	.575**	-.120
Triadic movement	-.664**	.430
Mean contour periodicity	.578**	.516*

* $p<.05$, ** $p<.01$

DISCUSSION

The results of this study provided an example of how judgments of similarity are dependent on the context. A re-analysis of a previous study (Rosner & Meyer, 1986) as well as results by Lamont and Dibben (2001) and others (McAdams *et al.*, 2004) suggested that listeners use the most salient variation between stimuli as the deciding factor in similarity judgments. Although the contextual effects of musical similarity cannot be fully demonstrated in a single experiment, we made an effort to pay attention to its importance by presenting examples from the literature, and by reanalyzing one study in terms of context effects in similarity formation (Rosner & Meyer, 1986). In our experiment, we aimed to provide a well-defined context for a melodic similarity rating task. This was achieved by choosing a representative sample of melodic phrases from a folk song collection, thus eliminating the issues of segmentation and supplying the listeners with a systematically varied set of stimuli. Listener's similarity ratings were predicted with geometric contour, pitch and interval content and contour periodicity vector differences as well as global predictors encapsulating the aforementioned measures. A fair amount (51 %) of participant similarity ratings were accounted for by the models and the multidimensional scaling solution of the rating data was highly explicable by the global predictors.

We believe that these results are in part demonstrating the salience of well-known similarity features (contour and pitch content) and in part the effects of the context. In other words, similarity is not just a distance between two objects but rather the context provides the framework within which specific features take on more or less importance. Therefore the results are indicative of the similarity within the particular set of stimuli used but can be rather different in different contexts. In this case, however, melodic phrases represented a large collection of typical phrases in folk songs and therefore it could argued that the results are suggestive of more general aspects of similarity perception in common-practice music. It also seems that the global predictors reflect the variance within the stimuli set in a more straightforward manner than the pair-wise predictors.

In an ideal similarity model, the features that contribute to similarity would be dynamically modified by the salient variation within the context of comparison. Although this operation has been carried out in recent computational work on similarity in other domains by applying principal components analysis to the data set (see *e.g.*, Tredoux, 2002 for analysis of facial similarity), this is seldom feasible in practical applications of melodic similarity. For example, a person who wishes to find a tune from a large digital library of music is not aware of the salient variation in the various dimensions the archive possesses. Moreover, a query-mechanism embedded into such a digital library has no knowledge of the context in which the user envisions the search query. An iterative process where the person redefines the parameters of the query with the feedback from the previous search might capture some aspects of the context. Nevertheless, the effect of context may be better dealt with by paying more attention to the global variation within the similarity set.

Address for correspondence:
Tuomas Eerola
Department of Music
University of Jyväskylä, Finland
P.O.Box 35, FI-40014 University of Jyväskylä
e-mail: tuomas.eerola@campus.jyu.fi

Micah Bregman
Department of Music
University of Jyväskylä, Finland
P.O.Box 35, FI-40014 University of Jyväskylä
e-mail: micah.bregman@gmail.com

• REFERENCES

Adams, C. (1976). Melodic contour typology. *Ethnomusicology*, *20* (2), 179-215.

Ahlbäck, S. (1997). A computer-aided method of analysis of melodic segmentation in monophonic melodies. In A. Gabrielsson (ed.), *Proceedings of the Third Triennial ESCOM Conference* (pp. 263-68). Uppsala, Sweden.

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

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

Brust, J. C. M. (2003). Music and the neurologist: A historical perspective. In R. Zatorre & I. Peretz (eds), *The cognitive neuroscience of music* (pp. 181-91). London: Oxford University Press

Cambouropoulos, E. (2001). Melodic cue abstraction, similarity, and category formation: A formal model. *Music Perception*, *18 (3)*, 347-70.

Cambouropoulos, E., & Widmer, G. (2000). Automated motivic analysis via melodic clustering. *Journal of New Music Research*, *29* (4), 303-17.

Chiappe, P., U., & Schmuckler, M. A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review*, *4* (2), 254-59.

Deliège, I. (1996). Cue abstraction as a component of categorisation processes in music listening. *Psychology of Music*, *24* (2), 131-56.

Deliège, I. (2001a). Introduction: Similarity perception<->categorization<->cue abstraction. *Music Pereption*, *18* (3), 233-43.

Deliège, I. (2001b). Prototype effects in music listening: An empirical approach to the notion of imprint. *Music Perception*, *18 (3)*, 371-407.

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Perception*, *12 (1)*, 1-25.

Dowling, W. J. (1971). Recognition of inversions of melodies and melodic contours. *Perception & Psychophysics*, *9*, 348-49.

Dowling, W. J. (1994). Melodic contour in hearing and remembering melodies. In R. Aiello & J. A. Sloboda (eds), *Musical Perceptions* (pp. 173-90). New York: Oxford University Press

Downie, J. S. (2003). Music information retrieval. *Annual Review of Information Science and Technology*, *37*, 295-340.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception*, *18* (3), 275-96.

Eerola, T., & Toiviainen, P. (2004). *MIDI toolbox: MATLAB tools for music research*. Jyväskylä, Finland: University of Jyväskylä.

Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition*, *25* (2), 237-55.

Hewlett, W. B., & Selfridge-Field, E. (1998). *Melodic similarity: Concepts, procedures and applications* (Vol. 11). Cambridge, MA: MIT Press.

von Hippel, P. (2000). Questioning a melodic archetype: do listeners use gap-fill to classify melodies? *Music Perception*, *18* (2), 139-53.

von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception*, *18* (1), 59-85.

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics*, *24*, 551-65.

Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, *27*, 75-97.

Kohonen, T. (1997). *Self-Organizing Maps* (2nd ed.). Berlin: Springer.

Koniari, D., Predazzer, S., & Melen, M. (2001). Categorization and schematization processes used in music perception by 10- to 11-year-old children. *Music Perception*, *18* (3), 297-324.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage Publications.

Lamont, A., & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception*, *18* (3), 245-74.

Lemström, K., & Ukkonen, E. (2000). Including interval encoding into edit distance based music comparison and retrieval. In *Proc. AISB'2000 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science* (pp. 53-60). Birmingham, United Kingdom: AISB.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Marvin, E. W., & Laprade, P. A. (1987). Relating musical contours: extensions of a theory for contour. *Journal of Music Theory*, *31* (2), 225-67.

Massaro, D. W., Kallman, H. J., & Kelly, J. L. (1980). The role of tone height, melodic contour, and tone chroma in melody recognition. *Journal of Experimental Psychology: Human Learning & Memory*, *6* (1), 77-90.

McAdams, S., Vieillard, S., Houix, O., & Reynolds, R. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Perception*, *22*, 207-37.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review 100*, 254-78.

Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception*, *3* (1), 1-32.

Müllensiefen, D., & Frieler, K. (2004). Optimizing measures of melodic similarity for the exploration of a large folk song database. In *Proceedings of the 5th International Conference on Music Information Retrieval: ISMIR 2004* (pp. 274-80). Barcelona, Spain: Universitat Pompeu Fabra.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: the implication-realization model*. Chicago: University of Chicago Press.

Ó Maidín, D. (1998). A geometrical algorithm for melodic difference. In W. B. Hewlitt & E. Selfridge-Field (eds.), *Melodic similarity: Concepts, procedures and applications* (Vol. 11, pp. 65-72). Cambridge, MA: MIT press

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythm. *Music Perception*, *11* (4), 409-64.

Polansky, L. (1996). Morphological metrics. *Journal of New Music Research*, *25* (4), 289-368.

Quinn, I. (1997). Fuzzy extensions to the theory of contour. *Music Theory Spectrum*, *19* (2), 232-63.

Quinn, I. (1999). The combinatorial model of pitch contour. *Music Perception*, *16* (4), 439-56.

Rips, L. J. (1989). Similarity, typicality and categorization. In S. Vosniadou & A. Ortony (eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.

Rosner, B. S., & Meyer, L. B. (1982). Melodic processes and the perception of music. In D. Deutsch (ed.), *The Psychology of Music* (pp. 317-41). New York: Academic Press.

Rosner, B. S., & Meyer, L. B. (1986). The perceptual roles of melodic process, contour, and form. *Music Perception*, *4*, 1-40.

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

Schmuckler, M. A. (1999). Testing models of melodic contour similarity. *Music Perception*, *16* (3), 295-326.

Schoenberg, A. (1974). *Style & Idea*. London: Faber & Faber.

Schulkind, M. D., Posner, R. J., & Rubin, D. C. (2003). Musical features that facilitate melody identification: How do you know it's "Your" song when they finally play it? *Music Perception*, *21(2)*, 217-49.

Serafine, M. L., Glassman, N., & Overbeeke, C. (1989). The cognitive reality of hierarchic structure in music. *Music Perception*, *6*, 347-430.

Snyder, J. L. (1990). Entropy as a measure of musical style: The influence of a priori assumptions. *Music Theory Spectrum*, *12*, 121-60.

Tenney, J., & Polansky, L. (1980). Temporal Gestalt perception in music. *Journal of Music Theory*, *24* (2), 205-41.

Tervaniemi, M., Rytkönen, M., Schroger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory*, *8* (5), 295-300.

Toiviainen, P., & Eerola, T. (2002). A computational model of melodic similarity based on multiple representations and self-organizing maps. In C. Stevens, D. Burnham, G. McPherson, E. Schubert & J. Renwick (eds.), *7th International Conference on Music Perception and Cognition* (pp. 236-39). Sydney, Australia: Causal Productions.

Trainor, L. J., Desjardins, R. N., & Rockel, C. (1999). A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Australian Journal of Pyschology*, *51* (3), 147-53.

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied*, *8 (3)*, 180-93.

Tsogo, L., Masson, M. H., & Bardot, A. (2000). Multidimensional scaling methods for many-object sets: A review. *Multivariate Behavioral Research*, *35* (3), 307-19.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-52.

Uitdenbogerd, A., & Zobel, J. (1999). Melodic matching techniques for large music databases. In *International Multimedia Conference Proceedings of the seventh ACM international conference on Multimedia* (pp. 57-66). NY, USA: ACM Press.

White, B. (1960). Recognition of distorted melodies. *American Journal of Psychology*, *73*, 100-07.

Wust, O., & Celma, O. (2004). An MPEG-7 database system and application for content-based management and retrieval of music. In *Proceedings of the 5th International Conference on Music Information Retrieval: ISMIR 2004* (pp. 48-51). Barcelona, Spain: Universitat Pompeu Fabra.

Zbikowski, L. M. (1999). Musical coherence, motive, and categorization. *Music Perception*, *17* (1), 5-42.

- **Similitud melódica y contextual de frases de canciones folklóricas**

Se proponen varios modelos de similitud melódica y se juzgan mediante experimentos perceptivos. Se han preferido las variables de perfil melódico y altura, y también se han seleccionado variables músico-teóricas y estadísticas para explicar los índices de similitud. Un re-análisis de la obra temprana de Rosner y Meyer (1986) sugiere que factores de simple contextualización pueden ser también explicados con estímulos más complejos. Un nuevo experimento desarrollado sobre frases melódicas breves ha investigado la efectividad de las variables globales y las comparativas. Una solución graduada multidimensional indicó que tanto la dirección melódica como la altura son altamente relevantes para realizar tales juicios de similitud y que los aspectos más destacados de la melodía para llevar a cabo juicios de similitud son hechos relativamente simples, que dependen del contexto.

- **Similarità melodica e contestuale di frasi del canto popolare**

Svariati modelli di similarità melodica sono stati proposti e valutati attraverso esperimenti percettivi. Per spiegare le valutazioni di similarità si sono favorite variabili di contorno e contenuto di altezze, pur chiamando in causa anche variabili teorico-musicali e statistiche. Una nuova analisi di un precedente lavoro di Rosner e Meyer (1986) suggerisce che semplici aspetti contestuali possono anche essere assai esplicativi con stimoli più complessi. Un nuovo esperimento contenente brevi frasi melodiche indagava l'efficacia di alcune variabili globali e comparative. Una soluzione scalare multidimensionale ha indicato che sia la direzione melodica sia l'estensione delle altezze sono assai rilevanti per operare giudizi di similarità, e che nell'effettuare tali giudizi gli aspetti più salienti della melodia sono elementi relativamente semplici e dipendenti dal contesto.

- **Similarité mélodique et contextuelle de phrases tirées de chants folkloriques**

On a proposé différents modèles de similarité mélodique qui ont été évalués par des expériences sur la perception. Les variables de hauteur et de contour ont été utilisées de préférence bien que des variables musicales théoriques et statistiques aient aussi été invoquées pour expliquer les classifications de similarité. Une nouvelle analyse des anciens travaux de Rosner et Meyer (1986) semble montrer que de simples caractéristiques contextuelles peuvent aussi expliquer la chose avec des stimuli plus complexes. Une nouvelle expérience a été entreprise utilisant de courtes phrases mélodiques pour étudier l'efficacité de plusieurs variables comparatives générales. Une solution d'échelle pluridimensionnelle a montré que la direction de la mélodie et le registre de hauteurs sont extrêmement pertinents dans la formation de jugements de similarité et que les aspects les plus saillants de la mélodie sont des caractéristiques assez simples et dépendantes du contexte.

• **Melodische und kontextuelle Ähnlichkeit von Volksliedphrasen**

Verschiedene Modelle der melodischen Ähnlichkeit wurden vorgeschlagen und in perzeptuellen Experimenten überprüft. Dabei wurden Variablen der Kontur und der Tonhöhe bevorzugt, obwohl musiktheoretische und statistische Variablen ebenso Ähnlichkeitsbeurteilungen erklären können, wie behauptet wurde. Eine Re-Analyse der Arbeit von Rosner und Meyer (1986) verweist darauf, dass einfache kontextuelle Merkmale auch bei komplexeren Stimuli von hohem Erklärungswert sein können. Mit einem neuen Experiment, das kurze melodische Phrasen enthält, wurde die Effektivität verschiedener globaler und komparativer Variablen untersucht. Eine multidimensionale Skalierungslösung zeigt, dass sowohl die melodische Richtung als auch der Tonhöhenbereich hochrelevant für solche Ähnlichkeitsbeurteilungen sind, und dass die auffälligsten Aspekte der Melodie in der Beurteilung von Ähnlichkeit relativ einfache kontextabhängige Merkmale sind.