*Music Perception*
Spring 2001, Vol. 18, No. 3, 275–296


# Statistical Features and Perceived Similarity of Folk Melodies

TUOMAS EEROLA, TOPI JÄRVINEN,
JUKKA LOUHIVUORI, & PETRI TOIVIAINEN
*University of Jyväskylä*

Listeners are sensitive to pitch distributional information in music (N. Oram & L. L. Cuddy, 1995; C. L. Krumhansl, J. Louhivuori, P. Toiviainen, T. Järvinen, & T. Eerola, 1999). However, it is uncertain whether frequency-based musical features are sufficient to explain the similarity judgments that underlie listeners' classification processes. A similarity rating experiment was designed to determine the effectiveness of these features in predicting listeners' similarity ratings. The material consisted of 15 melodies representing five folk music styles. A multiple regression analysis showed that the similarity of frequency-based musical properties could account for a moderate amount (40%) of listeners' similarity ratings. A slightly better predictive rate (55%) was achieved by using descriptive variables such as number of tones, rhythmic variability, and melodic predictability. The results suggest that both measures were able to capture some aspects of the structures that portray common salient dimensions to which listeners pay attention while categorizing melodies.

Aikaisemmissa tutkimuksissa on osoitettu, että musiikin tilastollisilla tapahtumilla, kuten sävelten määrillä ja tyypillisillä intervalleilla, on merkitystä, kun kuulijat muodostavat käsityksiään musiikin rakenteesta (N. Oram & L. L. Cuddy, 1995; C. L. Krumhansl, J. Louhivuori, P. Toiviainen, T. Järvinen, & T. Eerola, 1999). Näiden piirteiden voidaan olettaa olevan tärkeitä myös musiikin luokittelussa. Toistaiseksi ei kuitenkaan tiedetä, miten hyvin tilastollisilla piirteillä voitaisiin musiikin luokittelua selittää. Tätä testattiin kuulijoille järjestetyn samanlaisuus-arviointitehtävän avulla. Tutkimuksen materiaali koostui 15 melodiasta, jotka edustivat viittä eri kansanmusiikkityyliä. Regressioanalyysi paljasti, että musiikin tilastollisten piirteiden samanlaisuus pystyi selittämään kohtuullisen määrän (40%) kuulijoiden antamista samanlaisuus-arvioista. Hieman parempi selitysaste (55%) saavutettiin kuvaavilla muuttujilla, joita olivat melodian laajuus ja ennakoitavuus sekä rytmin vaihtelevuus. Näin ollen tulokset antavat aiheen olettaa, että musiikin tilastolliset piirteet ja kuvailevat muuttujat vaikuttavat kuulijoiden luokittelupäätöksiin.

Address correspondence to Tuomas Eerola, Department of Music, P. O. Box 35, FIN-40351 Jyväskylä, Finland. (e-mail: ptee@cc.jyu.fi)
ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

The ability to classify musical styles is an important and intriguing task from the perspective of music cognition. This process, which listeners usually do effortlessly, involves the integration of a number of perceptual processes. Recent summaries on categorization divide these processes into two groups: (1) rule application and (2) similarity computations (Hahn & Chater, 1998; Smith, Patalano, & Jonides, 1998). In this article, we consider similarity computations using the statistical frequencies of events, which have been shown to be influential in learning and perception of language and sound patterns (e.g., Saffran, Johnson, Aslin, & Newport, 1999). The line of inquiry is limited to melodic similarity, because this approach allows one to test and develop the frequency-based measures of melodic similarity that can be used to tackle some of the categorization and classification challenges that music history holds for us. Melodies drawn from folk music are suitable materials because they are relatively simple, monophonic yet realistic music. The authors have also had considerable previous experience with this kind of stimuli (Krumhansl, Louhivuori, Toiviainen, Järvinen, & Eerola, 1999; Krumhansl, Toivanen, Eerola, Toiviainen, Järvinen, & Louhivuori, 2000).

## MELODIC SIMILARITY

A growing number of studies have examined melodic similarity, and a number of experiments have shed light on different aspects of this phenomenon. Findings by Dowling (1971, 1978) indicate that one of the main factors of similarity is contour information, which is essential in short-term comparisons (Dowling & Bartlett, 1981) and in shorter melodies (Carterette, Kohl, & Pitt, 1986; Cuddy, Cohen, & Mewhort, 1981; Edworthy, 1985). Some studies have concentrated on melodic archetypes (Rosner & Meyer, 1982, 1986), hierarchical structure (Serafine, Glassman, & Overbeeke, 1989), themes (Pollard-Gott, 1983), motifs (Deliège, 1996; Lamont & Dibben, 1997; Zbikowski, 1999), whether melodies use scalar or nonscalar tones (Bartlett & Dowling, 1980, 1988; Dowling & Bartlett, 1981). More recently, studies have been focused on transposed melodies (Van Egmond, Povel, & Maris, 1996), the effects of pitch direction, contour, and pitch information (Dewitt & Crowder, 1986; Eiting, 1984; Freedman, 1999; Hofmann-Engl & Parncutt, 1998; Quinn, 1999; Schmuckler, 1999), and pitch range and key distance (Van Egmond & Povel, 1996). Commonly, rhythm has been considered as a separate entity (Gabrielsson, 1973; Palmer & Krumhansl, 1990; Simpson & Huron 1993) except by Monahan and Carterette (1985), who studied both rhythm and tonal dimensions as constituents of similarity. In this issue, there are other recent additions to melodic similarity such as cue abstraction (Deliège, 2001; see also Deliège, 1996, 1997), explorative search for the components of similarity in tonal and atonal music (Lamont & Dibben, 2001), and also development of similarity judgments during childhood (Koniari, Predazzer ,& Mélen, 2001) and infancy (Mélen & Wachsmann, 2001).

Theoretical models of melodic similarity include Cambouropoulos' (1997, 2001; Cambouropoulos & Smaill, 1995) formal definition of similarity based on the number of coinciding attributes of melodies. Anagnostopoulou and Smaill (2000) consider similarity as sets of properties on different hierarchical levels, the properties ranging from pitch-class sets to tempo and dynamic descriptions of atonal music. Discussion about theoretical models of melodic similarity was recently supplemented by Hewlett and Selfridge-Field's (1998) collection of articles as a volume of *Computing in Musicology*. Smith, McNab, and Witten (1998; also Orpen & Huron, 1992) defined similarity as the complexity of the transformation process involved in mapping one object onto the other. Cope's (1991, 1998) solution for categorizing music is to distinguish individual, characteristic signatures for specific composers. Models that deal with contour and interval information of the melodies include work by Deutsch and Feroe (1981), Ó Maidín (1998), and Hofmann-Engl and Parncutt (1998). The wide range of the focus of the research and the models can be credited to the multidimensional nature of melodic similarity.

The approach used in this article is different from the previous approaches in the following ways. First, the statistical properties of the melodies are hypothesized to provide perceptually salient cues for similarity judgments; second, the degree of match between listeners' similarity judgments and the similarity of statistical properties of melodies is examined in an experiment; and finally, a number of other plausible predictors for overall similarity of the melodies are investigated.

## SIMILARITY AND STATISTICAL PROPERTIES OF THE MELODIES

Earlier, category formation research held that the frequency of events and features is prominent in categorization. That was the notion behind Tversky and Kahneman's (1973) availability heuristic, where more frequent events become better established in memory. This notion is also present in the influential theory by Rosch (Rosch & Lloyd, 1978). The current research paradigm in category formation and representation still maintains that frequent stimuli have a greater impact on categorization than infrequent stimuli and that the frequencies of features are important components of similarity. It should also be remembered that categorization involves other processes as well (Barsalou, 1985; Barsalou, Huttenlocher, & Lamberts, 1998; Nosofsky, 1988, 1991). For instance, nonexemplar models (relying on rule-based processes) use high-level schemata to impose interpretations for the low-level representations (Hahn & Chater, 1998; Keil, Smith, Simons, & Levin, 1998). The influence of frequency information for similarity and category formation has mostly been studied in the field of language processing (Bassok & Medin, 1997; Trueswell, 1996). For example, Saffran and her colleagues (Saffran, Newport, & Aslin, 1996; Saffran et al., 1999) demonstrated how both adults and infants segmented "tone streams" according to their statistical properties, emphasizing the importance of frequency information in cognitive processes.

Research on music cognition and learning has demonstrated the effect of statistical information for learning and perception by means of both cross-cultural studies, using for example North Indian (Castellano, Bharucha, & Krumhansl, 1984), Balinese (Kessler, Hansen, & Shepard, 1984), and North Sami music (Krumhansl et al., 2000), as well as studies using melodies in which the statistical properties of music have been intentionally manipulated (Oram & Cuddy, 1995). These results show that listeners are sensitive to pitch distributional information. More specifically, for inexperienced listeners, the pitch distribution information presents cues concerning the basic melodic and tonal structure of the music, whereas for experienced listeners it gives rise to style-specific expectations. Also, composers and improvisers emphasize the important tones of the tonality by using them more frequently, with longer durations, and in the majority of case at structurally more important places than other tones (see Järvinen, 1995; Krumhansl, 1990; Knopoff & Hutchinson, 1983). In light of this evidence, it seems that statistical properties of melodies could provide a means for classification of musical styles in terms of their perceptual similarity. Indeed, studies using this approach have been appealing, for example, Järvinen, Toiviainen, and Louhivuori (1999) classified 10 different musical styles on the basis of the distributions of tones and tone transitions. The results, which were visualized by self-organizing maps (SOM), conformed with the musicological descriptions of the particular musical styles. Related methodology has been used by others (e.g., Atalay & Placek, 1997; Hörnel, 1998; Smaill & Westhead, 1993) with success. It is worth mentioning that the predecessors of these methods in music were conceived in ethnomusicology, where musical styles were classified according to the statistical distribution of different intervals, rhythmic patterns, or pitches (Freeman & Merriam, 1956; Lomax, 1968).

Despite the attractive results of the classifications based on statistical properties of melodies, it remains uncertain how well these methods simulate the human classification process. More specifically, it is not clear whether the properties that are used in such classifications of music are perceptually relevant and robust. To address this question, a similarity rating task was designed to investigate how effectively the statistical properties of the melodies can account for listeners' similarity judgments. In other words, listeners' similarity ratings were compared with the degrees of similarity of the statistical properties of the melodies.

# Methods

### SUBJECTS

Seventeen undergraduate music students (mean age = 23.4 years, *SD* = 4.2 years) participated in the study. They reported having studied music for a mean of 5.8 years (*SD* = 4.4 years) and having music as a hobby for a mean of 14.6 years (*SD* = 2.6 years). Three listeners reported being familiar with the particular melodies used in the experiment. Two of these responses concerned the Greek melody C2 (see Appendix), which had been used in a movie ("Never on Sunday") in the early 1960s. However, the majority of the participants were too young to be familiar with this instance of the song, a fact that was also apparent in their responses regarding their familiarity with the melodies used in the experiment.

### APPARATUS

The stimuli were generated using Sibelius software on an IBM-compatible 586-MHz computer with a Soundblaster AWE32 soundcard. The timbre used was English horn to ensure compatibility with previous experiments using similar folk melodies (Krumhansl et al., 1999, 2000). In the experiment, the stimuli were played by the computer using MEDS software (Kendall, 1999), fed through a Mackie CR1604-VL2 mixer/amplifier, and presented via two Yamaha active MSP5 speakers at a comfortable listening level.

### STIMULUS MATERIALS

Melodies from five distinct folk music styles were selected for the experiment. These were North Sami yoiks (Y), Finnish Spiritual folk hymns (H), Irish hornpipes (I), German folksongs (G), and Greek folksongs (C). Folk melodies were selected first for their musical validity and simplicity; second because it was assumed that melodies from different national styles would possess the natural variation of differences within melodies that is necessary for a similarity experiment. Three typical melodies of each style were chosen by native experts (see Appendix). The selection criteria included typicality, major mode, and moderate tempo. All 15 resulting monophonic melodies were approximately equal in length (mean duration = 17 s, *SD* = 1.3 s) but did not necessarily include the entire song, although all contained complete phrases. The melodies were transcribed from the original scores, transposed to C major, encoded as MIDI files, and had the same tempo (110 bpm), timbre (English horn), and velocity of tones. The playback was controlled by a computer. Because of length restrictions, only a few sequences could be used in the present experiment, and thus the results are not representative as far as stylistic categorization. In Rosch's terms, stylistic categorization is situated at a superordinate level that can be achieved only on the basis of a large number of examples at the basic level (Rosch & Lloyd, 1978). This limitation, however, plays no direct role here as our primary interest is on similarity formation rather than categorization.

### PROCEDURE

Subjects were given the instructions in written form, and the instructions were also verbally explained. They were told that their task was to rate the similarity of pairs of melodies on a scale from 1 to 9, where 1 corresponded to "very similar" and 9 to "very dissimilar." The subjects made the ratings by using a slider on the computer screen. All 105 possible combinations were randomly paired and ordered. The interstimulus duration between melodies was 1500 ms, which consisted of a 600-ms pause after the end of the first stimulus, followed by a 300-ms sine wave ($C_7$, 2093 Hz) beep indicating the end of the first stimulus and the beginning of the second stimulus. After the beep, there was another 600-ms pause before the beginning of the second stimulus. All subjects were tested individually in a sound-isolated room. Before the actual experiment, the subjects filled in a musical background questionnaire. Having read the instructions concerning the experiment, the subjects did three practice trials with materials that were not used in the experimental trials. After the experimental trials, the subjects indicated their previous familiarity with the melodies used in the experiment and were debriefed of the purpose of the experiment. The total duration of the experiment with the instructions, practice trials, and the survey of familiarity with the melodies was about 75 min.

## SIMILARITY MEASURES FOR THE MELODIES

The first set of similarity measures for the melodies was derived from the *statistical properties* of the melodies. These frequency-based statistical variables were the *distribution of the tones, intervals,* and *tone durations,* as well as the *distribution of two-tone transitions, interval transitions,* and *duration transitions.* Three sets of statistical variables were created from these and obtained for each melody (see the summary of the variables in Table 1). The first set was based on the raw frequencies of tones in the melodies.

The second set of similarity measures based on statistical properties of melodies was obtained by weighting all tones in the melody according to their duration. This weighting was done because longer tones are perceptually more salient than shorter tones (see, e.g., Boltz, 1993; Castellano, Bharucha, & Krumhansl, 1984; Krumhansl, 1991; Monahan & Carterette, 1985). The weight assigned to a particular note duration was defined in terms of the multiples of 16th notes (the shortest duration in this material).

The third set was obtained by analyzing three metrically hierarchical levels. This analysis was done because metrical hierarchy has been shown to influence the perception of tones, tonality, and meter (Dibben, 1994; Palmer & Krumhansl, 1990, Schmuckler & Boltz, 1994; Serafine et al., 1989). It was assumed that incorporating several hierarchical levels into the analysis would allow us to determine the relative contribution of each level to the similarity formation. The three hierarchical levels used were quarter note, half note, and whole note levels, and the coding followed the structural patterns proposed by Lerdahl and Jackendoff (1983). In this scheme, for example in 4/4 time signature, the initial beat of each measure represents the highest hierarchy (here whole note level); the first and the third beat, the second highest level (half note level); and the first, second, third, and fourth beats, the third level (quarter level). (See similar quantification schemes in Serafine et al., 1989).

For each distribution, the degree of similarity between each pair of melodies was determined by using the city block distance between the particular distributions. In other words, the degree of similarity of a given distribution was calculated according to

$$\sum_i \left| a_i - b_i \right|, \tag{1}$$

where $a$ and $b$ denote the distributions obtained from the two melodies. Table 1 provides a description of the variables. City block distance was used instead of correlation coefficient because it retains the variance across the components. All the analysis procedures began with the MIDI representations of the melodies used in the experiment, which were converted into kern representation and analyzed with Humdrum Toolkit (Huron, 1994) and additional Perl and shell scripts.

The second set of similarity measures was based on a number of *descriptive* variables that were obtained for each melody. These were gathered because it was assumed that listeners' overall judgments of similarities of the melodies might also reflect important internalized representations of musical information. Also these overall variables might be useful in connecting listeners' responses to distinctive characteristics of the melodies. The descriptive variables consisted of three groups. The first of these variables used tonal hierarchy information (Krumhansl, 1990) and was called *tonal stability.* It was measured as the correlation between the tone profile of the melody and the tonal hierarchy profile (in this case, the C-major probe-tone profile). A second group of descriptive variables dealt with the qualities of successive intervals; these variables were derived from Narmour's implication-realization model (1990) of melodic continuations and coded according to Krumhansl (1995a). These variables were *mean proximity of tones, registral return, registral direction, closure, intervallic difference,* and *consonance.* The motivation for using these variables in

TABLE 1
**Statistical Measures Extracted from Melodies**

| Variable | Description | No. of Components |
|---|---|---|
| Distribution of tones | All chromatic tones in one octave | 12 |
| Distribution of intervals | P8, M7, m7, M6, m6, P5, d5, P4, M3, m3, M2, m2, both up and down plus P1 | 25 |
| Distribution of durations | 𝅝., 𝅝, 𝅗𝅥., 𝅗𝅥, ♩., 𝅗𝅥(3), ♩, ♩(3), ♪., ♪, 𝅘𝅥𝅯., 𝅘𝅥𝅯, ♪(3), 𝅘𝅥𝅯, 𝅘𝅥𝅯(3), 𝅘𝅥𝅰 | 15 |
| Distribution of two-tone transitions | One octave by two octaves | 300 |
| Distribution of interval transitions | Two octaves by two octaves | 625 |
| Distribution of duration transitions | 15 durations by 15 durations | 225 |

predicting listeners' ratings of similarity stems from studies by Krumhansl (1995b) and Schellenberg (1996, 1997), who noticed that melodies representing different styles (Chinese and British folk songs, and atonal songs) differed to some extent according to these variables. *Tessitura* (melodic range in semitones) and *mean pitch* were added to the analysis as variables because they are important in discriminating melodies (Andrews & Dowling, 1991; Deutsch, 1997; Van Egmond et al., 1996). The third group of variables represented the rhythmic qualities of melodies. These were *syncopation, rhythmic variability, rhythmic activity,* and the *total number of tones*. Syncopation is a measure of deviation from the anticipated, regular beat pattern (Drake, Dowling, & Palmer, 1991; Povel & Essens, 1985) and may be considered as a moment where metrical stress is absent. Details of the syncopation coding were obtained from Huron (1994, pp. 441–442). Rhythmic variability was defined as the standard deviation of the notated durations, and rhythmic activity was defined as the number of tones per second. Finally, an aggregate function of tonal stability, proximity, intervallic difference, syncopation, and rhythmic variability, called *melodic predictability*, was computed for each melody. This measure has been developed by Eerola and North (2001) to account for the overall melodic predictability. For each of the descriptive variables, the degrees of similarity between the melodies were calculated as the absolute differences between the values of the particular variable.

## Results

### INTERSUBJECT CORRELATIONS

The similarity ratings were consistent across subjects as the mean intersubject correlation was significant ($r = .41$, $df = 103$, $p < .001$). In other words, no effects of order, familiarity, or musical background emerged from this analysis, and thus the data from all listeners were pooled into a single group for analysis.

### PERCEIVED DISTANCES BETWEEN THE MELODIES

The mean similarity ratings of listeners were analyzed by using a multidimensional scaling (ALSCAL) algorithm. Starting from the distances between a set of items, this algorithm maps these items onto a low-dimensional space while attempting to preserve the interitem distances. The multidimensional scaling solution can be interpreted as displaying the salient dimensions that underlie the perceptual experience (see Nosofsky, 1992). The difference between the two-dimensional ($R^2$ = .81) and three-dimensional ($R^2$ = .89) solution was small. If we look at the two-dimensional solution (Figure 1), few clear patterns emerge; the Greek songs (C1, C2, and C3) are farther away from others, meaning that these melodies were most distinctly different from all the other melodies. The three yoiks (Y1, Y2, and Y3) were perceived as melodically homogeneous and are clustered closely together. It is interesting to note that the Finnish spiritual folk hymns (H1, H2, and H3) were fairly similar and that the listeners found one German folk song (G2) and one Irish folk song (I2) to bear a resemblance to hymns. This may have occurred because German folk melodies and Finnish spiritual folk hymns share a number of common elements and are music-historically related (Suojanen, 1984). As predicted by Rosch's theory of categorization, it is evident from the figure that three examples per style is

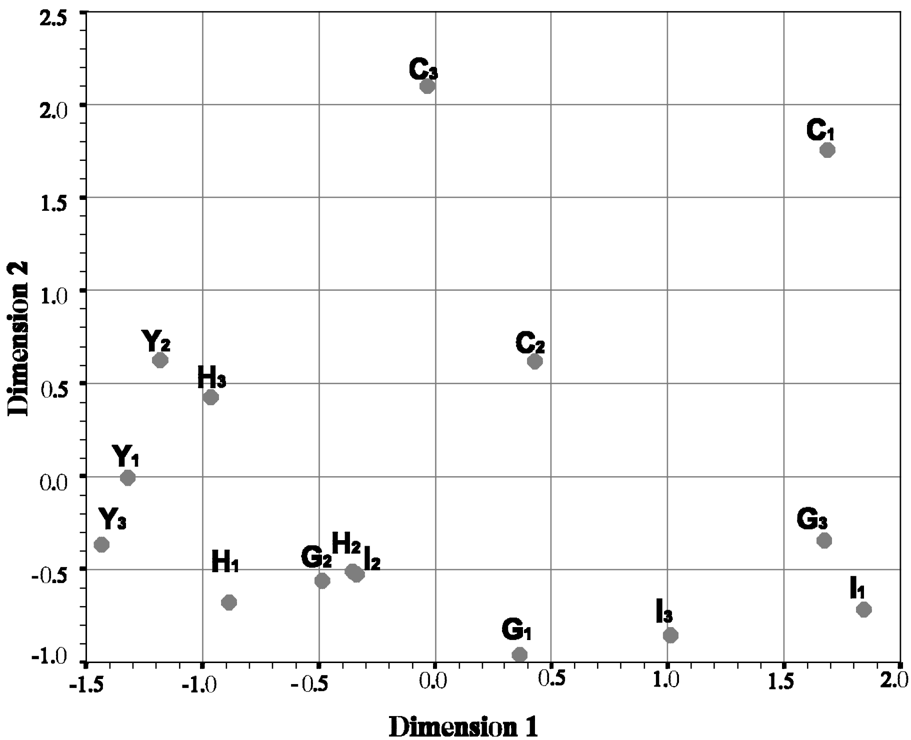



**Fig. 1.** Two-dimensional classical multidimensional scaling solution ($R^2$ = .81, $N$ = 15) for all the melodies. Y = Yoiks, H = Hymns, G = German, I = Irish, C = Greek.

not enough to create stylistic categories and that the results reflect more individual properties than style-specific qualities of the melodies. Even if we had used more examples of each style, we might not have obtained clear and distinct categories in the scaling solution because real musical examples drawn from a particular style are not necessarily truly homogeneous.

## ASSOCIATION BETWEEN THE SIMILARITY MEASURES DERIVED FROM THE STATISTICAL PROPERTIES OF THE MELODIES AND THE LISTENERS' SIMILARITY RATINGS

The similarity measures derived from the statistical properties of the melodies were compared with the listeners' similarity ratings by using a stepwise multiple regression analysis. The similarity measures were regressed upon the similarity ratings of the listeners for all pairs of melodies. The overall prediction rate of the unweighted frequency-based predictors was fairly low, ($R^2 = .39$, $F = 21.80$, $df = 3,101$, $p < .001$) and revealed that the distribution of duration transitions explained 20.1% of the variance in listeners' similarity ratings, and note transitions and duration distributions added 13.4% and 5.8%, respectively, to the prediction. As a result, it can be summarized that the melodies possessing similar rhythms and similar note transitions were judged to be similar by the listeners. The overall predictive power of the individual analyses and predictors is shown in Table 2. In order to portray the individual contribution of each predictor in the final regression equation, squared semipartial correlations are given in Table 2. In squared semipartial correlation, the effects of other variables have been controlled, which makes it a useful index of the importance of particular predictors in regression (Darlington, 1990; Howell, 1997). In the case of a regression model including only one predictor, the semipartial correlation is the same as the Pearson correlation coefficient.

Next, the variables that were weighted by the durations of the tones were analyzed by using the same procedure. Table 2 shows that these variables yielded the best predictive power. The variables that explained the variance were the same as those obtained in the original analysis except for the addition of the distribution of intervals. This suggests that incorporating a simple coding for the saliency of events in music increases the fit between the perceptual judgments and the theoretical similarity between the melodies.

The same analysis was performed separately for the three hierarchically encoded frequency-based sets of variables. This analysis attained a lower prediction rate ($R^2 = .22$) than the original encoding of the frequency-based variables and resulted in one significant predictor for each analysis (see Table 2). However, it is worth noting that the distribution of two-tone

TABLE 2
**Multiple Regression Results for Similarity of Frequency-Based Variables and Listeners' Similarity Ratings**

| Type of Similarity Measure | Prediction Rate |
|---|---|
| *All frequency-based variables* | $R^2$ = .39, $F$ = 21.8, $df$ = 3,101, $p$ < .001 |
| Distribution of tones | $sr^2$ = .04, n.s. |
| Distribution of intervals | $sr^2$ = .00, n.s. |
| Distribution of durations | $sr^2$ = .24* |
| Distribution of two-tone transitions | $sr^2$ = .37* |
| Distribution of interval transitions | $sr^2$ = .02, n.s. |
| Distribution of duration transitions | $sr^2$ = .36* |
| *Weighted by durations* | $R^2$ = .45, $F$ = 20.0, $df$ = 4,100, $p$ < .001 |
| Distribution of two-tone transitions | $sr^2$ = .29* |
| Distribution of duration transitions | $sr^2$ = .32* |
| Distribution of durations | $sr^2$ = .28* |
| Distribution of intervals | $sr^2$ = .19* |
| *Hierarchical measures, quarter note level* | $R^2$ = .22, $F$ = 29.0, $df$ = 1,103, $p$ < .001 |
| Distribution of two-tone transitions | $sr^2$ = .47* |
| *Hierarchical measures, half note level* | $R^2$ = .14, $F$ = 16.3, $df$ = 1,103, $p$ < .001 |
| Distribution of intervals | $sr^2$ = .37* |
| *Hierarchical measures, whole note level* | $R^2$ = .05, $F$ = 5.5, $df$ = 1,103, $p$ < .001 |
| Distribution of intervals | $sr^2$ = .23* |

Note—$sr^2$ is the squared semipartial correlation representing the unique proportion of variance explained by the relevant predictor.
*$p$ < .001.

transitions and the distribution of intervals become better predictors of similarity when the measures are based on reductions of the melodies. The results based on the quarter and half note levels imply that listeners may be forming their notion of similarity for the pairs of melodies from slightly abstracted versions of the melodies.

Finally, all three sets of frequency-based variables were entered in a single stepwise regression. This analysis obviously selected those variables that explained the greatest amount of variance in the previous analysis and the multiple $R$ of the resulting six predictor solution was .52 ($F$ =18.0, $df$ = 6,98, $p$ < .001). It might be concluded from this that although the predictive power of the frequency-based variables could be slightly improved by accounting for different hierarchical levels and relative durations of tones, these variables were only moderately successful in explaining listeners' similarity ratings.

Further analyses will be summarized briefly. A stepwise regression analysis was also performed between listeners' similarity ratings and similarities between the descriptive variables. The results (Table 3) indicate that six descriptive variables could explain as much as 62% of the variance in similarity ratings. These variables were, in order of importance, melodic predictability, mean pitch, tonal stability, consonance, number of tones, and

TABLE 3
**Multiple Regression Results for Similarity of Descriptive Variables and Listeners' Similarity Ratings**

| Type of Similarity Measure | Prediction Rate |
|---|---|
| *All descriptive variables* | $R^2 = .62$, $F = 26.5$, $df = 6{,}98$, $p < .001$ |
| Tonal stability | $sr^2 = .17$** |
| Mean proximity of tones | $sr^2 = .05$ |
| Registral return | $sr^2 = .04$ |
| Registral direction | $sr^2 = .11$ |
| Closure | $sr^2 = .14$* |
| Intervallic difference | $sr^2 = .19$ |
| Consonance | $sr^2 = .13$* |
| Tessitura | $sr^2 = .06$ |
| Mean pitch | $sr^2 = .33$*** |
| Syncopation | $sr^2 = .13$ |
| Rhythmic variability | $sr^2 = .10$ |
| Rhythmic activity | $sr^2 = .11$ |
| Total number of tones | $sr^2 = .19$** |
| Melodic predictability | $sr^2 = .51$*** |

*$p < .05$. **$p < .01$. ***$p < .001$.

closure. Note that the unique contribution of the weakest variables (closure, consonance, total number of tones) in the regression equation explain a similar amount of variance in similarity ratings as some of the variables that were not included in the regression equation (syncopation, intervallic difference). This indicates that it is difficult to assess the exact contribution of each variable without using a more controlled set of stimuli. Nonetheless, the descriptive variables were somewhat better predictors of melodic similarity than were frequency-based variables. At this point, however, it is difficult to assess whether these few descriptive variables better reflect the fundamental elements of melodic similarity than frequency-based variables.

The final regression analysis was conducted with all the frequency-based and descriptive variables. This yielded the best possible fit using these variables, accounting for almost 75% of the variance in listeners' similarity ratings ($R^2 = .74$, $F = 40.3$, $df = 7{,}97$, $p < .001$). Although this analysis suffers from potential "overfitting," it functions here as an estimate of a convergence between approaches. The degree of conformity also suggests that the most relevant variables of this analysis could together create a plausible categorization solution for the melodies if this was accomplished by using a suitable clustering algorithm. Still, categorization based on these variables should be viewed with caution before further experiments have assessed the extent to which the results may be generalized. Another way of looking at the pattern of responses is to investigate the dimensions that characterized listeners' similarity ratings. This will be addressed next.

## SALIENT DIMENSIONS OF THE MELODIES

As the connection between the statistical properties and the perceived similarities of the melodies was moderate, the salient dimensions of listeners' ratings were studied in more detail. This was achieved by correlating the dimensions of the scaling solution with the initial, overall measures of the descriptive variables for each melody. This correlation showed that Dimension 1 in the two-dimensional solution correlated with mean pitch, melodic predictability, registral return, registral direction, rhythmic activity, and tonal stability ($r$ = .92, .77, .61, .65, .58, –.52, respectively, $p < .05$ and $df = 13$ in all cases) and Dimension 2 correlated with syncopation, rhythmic variability, and proximity ($r$ =.55, .53, –.55, $df = 13$, $p < .05$). Overall, the same variables that predicted listeners' similarity ratings were evident in this analysis. Thus, there is no clear-cut explanation for Dimension 1, except that it is influenced by mean pitch height of melodies and could otherwise be interpreted as the predictability of melodies, which consists of regularity of large intervals and tonal stability. Furthermore, the listeners evidently attended to the rhythmic qualities of the melodies while making their similarity ratings as indicated by Dimension 2. Broadly speaking, pitch and rhythm could be said to be the most feasible categorization factors of these melodies, a concept that largely corresponds to the findings of Monahan and Carterette (1985). However, analysis of the dimensions of the scaling solution made it clear that these dimensions could not be easily interpreted because of the multidimensional nature of the melodies used in the experiment.

## DISCUSSION

Only moderate success was achieved when the similarities of the statistical properties of the melodies were used to explain the perceived similarities of the melodies. Although some evidence exists that frequencies of events may be useful as cues for similarity, the raw frequency counts of tones, intervals, and durations and their first-order transitions were not particularly effective predictors of melodic similarity. Several factors might have affected the moderate degree of fit attained in this study. First, the small quantity of available statistical data does not do justice to the statistical approach. In the previous studies, considerably larger samples were used (e.g., 100 melodies in Järvinen et al., 1999; 80 works in Atalay & Placek, 1997). Therefore the results of these previous studies might reflect more appropriately the musical style in general, whereas the responses here were more driven by the unique features of the melodies. In effect, these unique features probably caused the listeners to adopt similarity rating strategies that were not accommodated by the measurement models. Second, the melodies were fairly long and did not contain an equal number of tones, thus causing difficulties for the listeners and the similarity models. Third and most important, it is safe to assume that the events in a melody are not equally salient. A glimpse of this was seen when a simple weighting of the events was considered. When the metrical hierarchy information and the durations of tones were taken into account, the predictive power of the frequency-based variables increased from 39% to 52%. The prediction rate might have been higher if the salience of individual events had considered aspects such as melodic (Huron & Royal, 1996), harmonic (Boltz, 1998), or contour accents. Another improvement would be to segment the melodies into smaller phrases for the analysis (e.g., Anagnostopoulou & Smaill, 2000). This segmentation would take into account the salience of phrase boundaries, and the resulting segments would be more in line with the memory limitations (Deliège, 1989; Jones & Boltz, 1989; Krumhansl, 1996). Further methods that account for similarity are the contour and interval difference models (see reviews in Hofmann-Engl & Parncutt, 1998 and Schmuckler, 1999). A common difficulty with these models is that it is difficult to apply them if the melodies are not isochronous as was the case in this study.[1] However, omission of contour models from this study might not have affected the results to a great extent, as contour seems to play a more important role in nondiatonic than in diatonic sequences (Bartlett & Dowling, 1980; Dowling & Fujitani, 1971; Dowling & Harwood, 1986).

If one bears in mind the caveats in the present experiment, the overall explanatory power of the frequency-based variables can be regarded as reasonable. Results show that the contribution of frequency-based and descriptive variables to similarity formation can be assessed and these variables can explain a part of the perceived similarities between melodies.

## General Discussion

Listeners' similarity ratings of folk melodies were explained by the degrees of similarity in the statistical properties and descriptive variables of the melodies. The results suggest that statistical measures are able to capture some basic aspects of structures that portray common salient dimensions to which listeners pay attention while categorizing melodies. The descriptive variables were somewhat better at explaining listeners' similarity ratings, but further research is warranted before their individual roles in similarity formation can be assessed.

1. In a separate analysis, the Hoffman-Engl and Parncutt (1998) model called *normalized interval difference model* and its two alternative formulations were applied to the first and the last 12 tones of the melodies. These measures could account for only a small amount of variance in the data (between the first 12 tones 4.6% and between the first and last 12 tones 4.7%, for a total of $R^2 = .092$, $F = 5.2$, $df = 2,102$, $p < .05$).

One reason for the moderate degree of success of the statistical analysis lies in the multidimensional nature of the melodies and the oversimplified representation of melodic information. In future studies, the multidimensionality could be reduced by maintaining better control of the parameters. Other experimental paradigms should also be used. One way of imposing higher level structure on the frequency-based classification methods would be distinguishing and weighting the frequency information according to the perceptual prominence of events. As shown earlier, when a simple weighting scheme was used, the connection between perceptual and frequency-based similarity improved. Another reasonable explanation for the moderate predictive power of the frequency-based variables was offered by Keil et al. (1998, p. 107), when they explained why higher order tabulations of similarity often fail: "bottom-up statistical patterns do not always drive reasoning: we often use high-level schema to impose interpretations of statistical patterns." In other words, the kinds of rules that are actually applied in evaluating melodic similarity should be investigated in detail. For example, a crucial study would combine the open-ended investigation of the principles in similarity judgments (see Lamont & Dibben, 2001) with the breakdown of the musical sequences into separate variables. This approach would attempt to find correlates between these different approaches. By doing so, the gap would be bridged between music analytical variables, statistical properties of the melodies, and the underlying reasons for the melodic similarity given both by nonmusicians and musicians. Nevertheless, questions about the representation and the similarity of melodies remain central in music perception. We also believe that the use of several approaches found herein will be essential to future research into melodic similarity and categorization.[2]

## References

Anagnostopoulou, C., & Smaill, A. (2000). Similarity and categorisation in Boulez' Parenthese from the Third Piano Sonata: A formal analysis. *Proceedings of the 6th ICMPC,* Keele, U.K.

Andrews, M. W., & Dowling, W. J. (1991). The development of perception of interleaved melodies and control of auditory attention. *Music Perception, 8,* 349–368.

Atalay, B. V., & Placek, R. (1997). Machine versus human: Responding to the task of identifying eras for keyboard pieces. In A. Gabrielsson (Ed.), *Proceedings of the Third Triennial ESCOM conference* (pp. 521–526). Uppsala, Sweden: Uppsala University.

2. This research was supported by the Pythagoras Graduate School for Music and Sound Research (funded by the Finnish Ministry of Education) to Tuomas Eerola. It was also supported by a postdoctoral research grant from the Academy of Finland to Topi Järvinen. Petri Toiviainen was supported by a Fulbright fellowship and a grant from the Academy of Finland. We are grateful for our native experts, Francis Kiernan, Almut Meyer-Toivanen, Pekka Toivanen, Sini Järvelä, Terhi Nurmesjärvi-Skaniakos, and Georgios Skaniakos for finding and selecting the material for the experiment. We are also thankful to reviewers for their helpful suggestions on a previous version of this article.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629–654.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology, 36,* 203–272.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 501–515.

Bartlett, J. C., & Dowling, W. J. (1988). Scale structure and similarity of melodies. *Music Perception, 5,* 285–314.

Bassok, M., & Medin, D. L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language, 36*, 311–336.

Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception and Psychophysics, 53,* 585–600.

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception and Psychophysics, 60,* 1357–1373.

Cambouropoulos, E., & Smaill A. (1995). A computational model for the discovery of parallel melodic passages. In L. Finarelli & F. Regazzi (Eds.), *Proceedings of the XI Colloquium on Musical Informatics* (pp. 67–70). Bologna, Italy: Associazione di Informatica Musicale Italiana.

Cambouropoulos, E. (1997). The role of similarity in categorisation: Music as a case study. In A. Gabrielsson (Ed.), *Proceedings of the Third Triennial ESCOM conference* (pp. 533–538). Uppsala, Sweden: Uppsala University.

Cambouropoulos, E. (2001). Melodic cue abstraction, similarity and category formation: A computational approach. *Music Perception, 18*, 347–370.

Carterette, E. C., Kohl, D. V., & Pitt, M. A. (1986). Similarities among transformed melodies: The abstraction of invariance. *Music Perception, 3,* 393–410.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General, 113,* 394–412.

Cope, D. (1991). *Computers and musical style.* The Computer Music and Digital Audio Series, 6. Madison, WI: A-R Editions.

Cope, D. (1998). Signatures and earmarks: Computer recognition of patterns in music. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Melodic similarity: Concepts, procedures, and applications* (pp. 129–138). Cambridge, MA: MIT Press.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 869–883.

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Deliège, I. (1989). Approche perceptive de formes contemporaines. In S. McAdams & I. Deliège (Eds.), *La musique et les sciences cognitives* (pp. 305–326). Bruxelles, Belgium: Pierre Mardaga.

Deliège, I. (1996). Cue abstraction as a component of categorisation processes in music listening. *Psychology of Music, 24,* 131–156.

Deliège, I. (1997). Similarity in processes of categorisation: Imprint formation as a prototype effect in music listening. *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation* (pp. 59–65). Edinburgh: University of Edinburgh.

Deliège, I. (2001). Prototype effect in music listening: An empirical approach of the notion on imprint. *Music Perception, 18*, 371–407.

Deutsch, D. (1997). The tritone paradox: A link between music and speech. *Current Directions in Psychological Science, 6,* 174–180.

Deutsch, D. ,& Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88*, 503–522.

Dewitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception, 3*, 259–274.

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

Dowling, J. W. (1971). Recognition of inversions of melodies and melodic contours. *Perception & Psychophysics, 9,* 348–349.
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. *Journal of the Acoustical Society of America, 49,* 524–531.
Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* Orlando, FL: Academic Press.
Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85,* 341–354.
Drake, C., Dowling, W. J., & Palmer, C. (1991). Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Perception, 8,* 315–334.
Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception, 2,* 375–388.
Eerola, T., & North, A. C. (2001). Measuring melodic complexity: An expectancy-based model. Manuscript submitted for publication.
Eiting, M. H. (1984). Perceptual similarities between musical motifs. *Music Perception, 2,* 78–94.
Freedman, E. G. (1999). The role of diatonicism in the abstraction and representation of contour and interval information. *Music Perception, 16,* 365–387.
Freeman, L. C., & Merriam, A. P. (1956). Statistical classification in anthropology: An application to ethnomusicology. *American Anthropologist, 58,* 464–472.
Gabrielsson, A. (1973). Similarity ratings and dimension analysis of auditory rhythm patterns. I & II. *Scandinavian Journal of Psychology, 14,* 138–160, 161–176.
Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition, 65,* 197–230.
Hewlett, W. B., & Selfridge-Field, E. (Eds.). (1998). *Melodic similarity: Concepts, procedures, and applications.* Cambridge, MA: MIT Press.
Hofmann-Engl L., & Parncutt R. (1998). Computational modeling of melodic similarity judgments: Two experiments on isochronous melodic fragments [On-line]. Available: http://freespace.virgin.net/ludger.hofmann-engl/similarity.html
Hörnel, D. (1998). A multi-scale neural-network model for learning and reproducing choral variations. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Melodic similarity: Concepts, procedures, and applications* (pp. 141–157). Cambridge, MA: MIT Press.
Howell, D. C. (1997). *Statistical methods for psychology.* 4th ed. Fairfield, CT: Duxbury Press.
HSHL (1988). *Siionin virret ja halullisten sielujen hengelliset laulut* [Hymns of zion and spiritual songs for devoted souls]. Rauma: Rukoilevaisen Kansan Yhdistys.
Huron, D. (1994). *UNIX tools for musical research: The humdrum toolkit reference manual.* Stanford, CA: Center for Computer Assisted Research in Humanities.
Huron, D., & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception, 13,* 489–516.
Järvinen, T. (1995). Tonal hierarchies in jazz improvisation. *Music Perception, 12,* 415–437.
Järvinen, T., Toiviainen, P., & Louhivuori, J. (1999). Classification and categorization of musical styles with statistical analysis and self-organizing maps. In A. Patrizio, G. A. Wiggins, & H. Pain (Eds.), *Proceedings of the AISB'99 Symposium on Musical Creativity* (pp. 54–57). Edinburgh: The Society for the Study of Artificial Intelligence and Simulation of Behaviour.
Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459–491.
Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition, 65,* 103–135.
Kendall, R. A. (1999). *Music Experiment Development System: MEDS 2000 Software for Windows 95/2000* [computer software]. Los Angeles: UCLA.

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the west. *Music Perception, 2,* 131–165.
Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory, 27,* 75–97.
Koniari, D., Predazzer, S., & Mélen, M. (2001). Categorisation processes in music perception during childhood. *Music Perception, 18*, 297–324.
Krumhansl, C. L. (1990). *Cognitive foundation of musical pitch*. New York: Oxford University Press.
Krumhansl, C. L. (1991). Memory for musical surface. *Memory and Cognition, 19,* 401–411.
Krumhansl, C. L. (1995a). Effects of musical context on similarity and expectancy. *Systematische Musikwissenschaft, 3,* 211–250.
Krumhansl, C. L. (1995b). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum, 17,* 53–90.
Krumhansl, C. L. (1996). A perceptual analysis of Mozart's piano sonata K.282: Segmentation, tension, and musical ideas. *Music Perception, 13,* 401–432.
Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception, 17,* 151–196.
Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition, 76,* 13–58.
Lamont, A., & Dibben, N. (1997). Perceived similarity of musical motifs: An exploratory study. In M. Ramscar, U. Hahn, E. Cambouropoulos, & H. Pain (Eds.), *Proceedings of SimCat 1997: An Interdisciplinary Workshop on Similarity and Categorisation* (pp. 143–149). Edinburgh: Department of Artificial Intelligence, University of Edinburgh.
Lamont, A. & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception, 18*, 245–274.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
Lomax, A. (1968). *Folk song style and culture*. Washington, D.C.: American Association for the Advancement of Science.
Mélen, M., & Wachsmann, J. (2001). Categorization of musical structures by 6-to-10 month old infants. *Music Perception, 18,* 325–346.
Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception, 3,* 1–32.
Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.
Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.
Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 3–27.
Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43,* 25–54.
O Maidín, D. (1998). A geometrical algorithm for melodic difference. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Melodic similarity: Concepts, procedures, and applications* (pp. 65–72). Cambridge, MA: MIT Press.
O'Neill, E. R. (1988). *Seoltoireacht Ghear*. Dublin: Geal Linn.
Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research , 57,* 103–118.
Orpen, K., & Huron, D. (1992). The measurement of similarity in music: A quantitative approach for non-parametric representations. *Computers in Music Research, 4,* 1–44.
Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 728-741.

Pollard-Gott, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive Psychology, 15,* 66–94.

Povel, D. J., & Essens, P. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception and Psychophysics, 37,* 1–7.

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

Rosch, E., & Lloyd, B. B. (1978). *Cognition and categorization.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosner, B., & Meyer, L. B. (1982). Melodic process and the perception of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 317–342). New York: Academic Press.

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

Rölleke, H. (Ed.) (1993). *Das Volksliederbuch.* Köln: Kiepenhauer & Witsch.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition, 58,* 75–125.

Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception, 14,* 295–318.

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

Schmuckler, M. A., & Boltz, M. G. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception and Psychophysics, 56,* 313–325.

Serafine, M. L., Glassman, N., & Overbeeke, C. (1989). The cognitive reality of hierarchic structure in music. *Music Perception, 6,* 397–430.

Simpson, J., & Huron, D. (1993). The perception of rhythmic similarity: A test of a modified version of Johnson-Laird's theory. *Canadian Acoustics, 21,* 89–90.

Smaill, A., & Westhead, M. D. (1993). Automatic characterisation of musical style. *Music education: An artificial intelligence approach* (pp. 157–170). Berlin: Springer-Verlag.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65,* 167–196.

Smith, L. A., McNab, R. J., & Witten, I. H. (1998). Sequence-based melodic comparison: A dynamic-programming approach. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Melodic similarity: Concepts, procedures, and applications* (pp. 101–117). Cambridge, MA: MIT Press.

Suojanen, P. (1984). *Finnish folk hymn singing: Study in music anthropology* (Publication 11). Tampere, Finland: University of Tampere Institute for Folk Tradition.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35,* 566–585.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232.

Van Egmond, R., & Povel, D.-J. (1996). Perceived similarity of exact and inexact transpositions. *Acta Psychologica, 92,* 283–295.

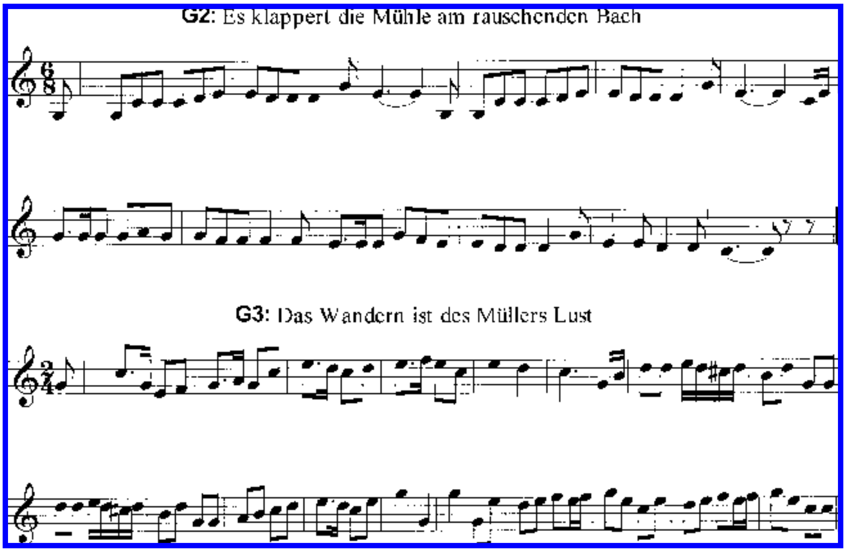
Van Egmond, R., Povel, D.-J., & Maris, M. (1996). The influence of height and key on the perceptual similarity of transposed melodies. *Perception and Psychophysics, 58,* 1252–1259.

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

## Appendix
## Melodies Used in the Experiment

Greek folk songs. (Collected and transcribed by Sini Järvelä.)

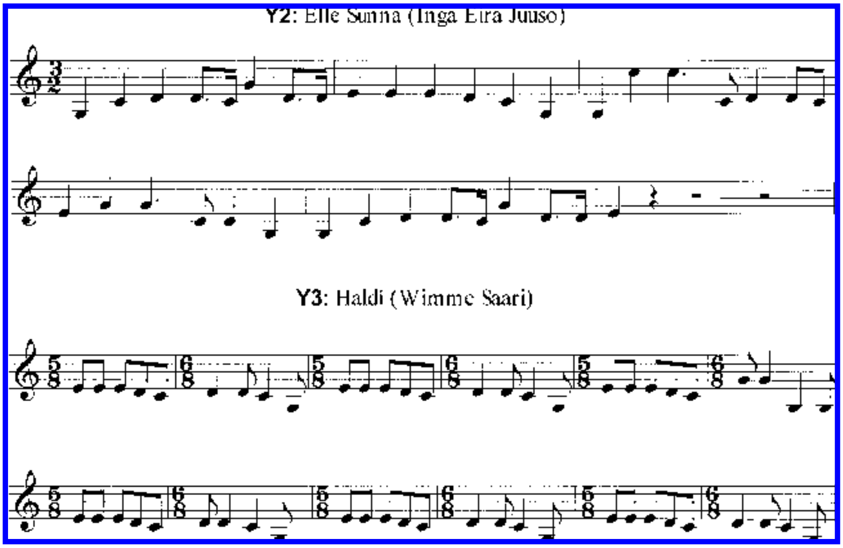

German folk songs. (From Rölleke, 1993.)




## Finnish spiritual folk hymns. (From "Halullisten Sielujen Hengelliset Laulut" [HSHL], 1998.)

IRISH HORNPIPES. (FROM O'NEILL, 1988.)


NORTH SAMI YOIKS. (TRANSCRIBED BY PEKKA TOIVANEN. FOR A DESCRIPTION OF THE YOIK CORPUS FROM WHICH THESE THREE MELODIES WERE CHOSEN, SEE KRUMHANSL ET AL., 2000. EACH YOIKER'S NAME IS INDICATED IN PARENTHESES.)


Y2: Elle Sunna (Inga Eira Juuso)
Y3: Haldi (Wimme Saari)