*Music Perception*
Winter 1999, Vol. 17, No. 2, 151–195



# Melodic Expectation in Finnish Spiritual Folk Hymns: Convergence of Statistical, Behavioral, and Computational Approaches

CAROL L. KRUMHANSL
*University of Jyväskylä and Cornell University*

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

This study of Finnish spiritual folk hymns combined three approaches to understanding melodic expectation. The first approach was a statistical style analysis of a representative corpus of 18 hymns, which determined the relative frequencies of tone onsets and two- and three-tone transitions. The second approach was a behavioral experiment in which listeners, either familiar (experts) or unfamiliar (nonexperts) with the hymns, made judgments about melodic continuations. The third approach simulated melodic expectation with neural network models of the self-organizing map (SOM) type (Kohonen, 1997). One model was trained on a corpus of Finnish folk songs and Lutheran hymns (Finnish SOM), while another was trained with the hymn contexts used in the experiment with the correct continuation tone (Hymn SOM). The three approaches converged on the following conclusions: (1) Listeners appear to be sensitive to the distributions of tones and tone transitions in music, (2) The nonexperts' responses more strongly reflected the general distribution of tones, whereas the experts' responses more strongly reflected the tone transitions and the correct continuations, (3) The SOMs produced results similar to listeners and also appeared sensitive to the distributions of tones and tone transitions, (4) The Hymn SOM correlated more strongly with the experts' judgments than the Finnish SOM, and (5) the principles of the implication-realization model (Narmour, 1990) were weighted similarly by the behavioral data and the Hymn SOM.

Tässä suomalaisia hengellisiä kansansävelmiä käsittelevässä tutkimuksessa pyrittiin selvittämään melodisia odotuksia kolmen tutkimusmenetelmän avulla. Ensimmäinen menetelmä oli kyseistä tyyliä edustavien 18 sävelmän tilastollinen analyysi, jossa määriteltiin sävelkorkeuksien sekä kahden ja kolmen sävelen siirtymien tilastolliset jakaumat. Toinen menetelmä oli behavioraalinen koe, jossa kuulijat arvioivat sävelmien jatkoja. Kuulijat

Address correspondence to Carol L. Krumhansl, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853 USA. (e-mail: clk4@cornell.edu)

jakaantuivat kahteen ryhmään: sävelmät tunteviin (asiantuntijoihin) ja sävelmiä tuntemattomiin (ei-asiantuntijoihin). Kolmannessa menetelmässä simuloitiin melodisia odotuksia itsejärjestäytyvään karttaan (Kohonen, 1997) perustuvalla keinotekoisella hermoverkkomallilla. Ensimmäiselle mallille opetettiin joukko suomalaisia kansanlauluja ja luterilaisia virsiä (suomalainen verkko), toiselle kokeessa käytettyjä hengellisiä kansansävelmiä (hengellinen verkko). Käytetyt menetelmät tuottivat yhteneviä tuloksia ja antoivat aihetta seuraaviin johtopäätöksiin: (1) kuulijat näyttävät olevan vastaanottavaisia musiikin säveljakaumille ja sävelsiirtymille, (2) ei-asiantuntijoiden vastaukset noudattivat enemmän sävelten yleistä jakaumaa, kun taas asiantuntijoiden vastaukset heijastivat enemmän sävelsiirtymiä ja sävelmien oikeita jatkoja, (3) hermoverkot tuottivat tuloksia, jotka olivat samankaltaisia kuulijoiden arvioiden kanssa ja jotka noudattivat sävelten ja sävelsiirtymien jakaumia, (4) hengellisen verkon tulokset korreloivat suomalaisen verkon tuloksia voimakkaammin asiantuntijoiden arvioiden kanssa, ja (5) behavioraaliset tulokset ja hengellinen verkko painottavat implikaatio-realisaatio-mallin (Narmour, 1990) periaatteita samalla tavalla.

Received July 13, 1998; revision received June 7, 1999; accepted for publication September 4, 1999.

The research reported in this article takes three approaches to studying melodic expectation in a style of Finnish music, spiritual folk hymns. The first, a statistical style analysis, determined tone distributions and common tone transitions. The second, a behavioral experiment, had listeners judge tones presented after short hymn fragments as to how well they fit with their expectations for melodic continuation. The third approach trained neural network models of the self-organizing map (SOM) type. This combination of approaches can be used to investigate the contributions to melodic expectation made by general principles of tone perception, schematic knowledge of the Western style, sensitivity to statistical regularities within the particular style, and precise knowledge of the individual hymns. Comparisons with the neural network model can assess its adequacy as a model of learning and uncover musical and perceptual features it abstracts through training.

Expectation is considered a basic component of music perception, operating on a variety of levels, including melodic, harmonic, metrical, and rhythmic. It is assumed to be important for the process of organizing sounds into coherent units, creating patterns of tension and relaxation that may play a role in generating musical emotions (Meyer, 1956). In the present context, we consider only tone-to-tone expectations in melodies. This focus was chosen because it has been the subject of a number of previous investigations (e.g., Carlsen, 1981; Carlsen, Divenyi, & Taylor, 1970; Cuddy & Lunney, 1995; Krumhansl, 1991, 1995a, 1995b, 1997; Schmuckler, 1989, 1990; Thompson, Cuddy & Plaus, 1997; Unyk & Carlsen, 1987), establishing experimental methods applicable to the question and providing comparative data. This focus was also chosen because the results can be used to test theoretical proposals concerning melodic expectation.

Bharucha (1987) makes the useful distinction between schematic expectation and veridical expectation. According to him, listeners abstract recurrent commonalities from the music they hear, forming schematic representations. These representations are assumed to be activated automatically when listening to music and guide expectations for events to follow. Examples of schematic expectation include key and harmonic relationships. In contrast, veridical expectation is generated either by the activation of memory traces for specific pieces or explicit prior knowledge about how the music continues. They may or may not be activated automatically. Schematic expectation may conflict with veridical expectation, with the result that a highly familiar piece can continue to be interesting (see also Meyer, 1967). Bharucha and collaborators (Bharucha, 1987, 1991, 1999; Bharucha & Olney, 1989; Bharucha & Todd, 1989) have developed models of musical expectation that are implemented as neural networks.

In the present study, the statistical style analysis considered a number of different measures. The relative frequency with which tones occur (independently of tone order) would be expected to reflect more schematic information about Western tonality (e.g., major and minor tonal hierarchies), whereas frequencies of transitions between tones (taking longer contexts into account) would be expected to capture more specific features of the style and the particular hymns. In the behavioral experiment, two groups of listeners participated. One group (nonexperts) had no previous exposure to the hymns and only a general familiarity with the style (although they were music students well-trained in Western music). The other group (experts) knew the particular hymns and were highly familiar with the style. We would expect the experts to show stronger veridical expectations than the nonexperts. Two self-organizing networks were trained. One (Hymn SOM) was trained with only the hymn excerpts used in the experiment with the correct continuation tones to simulate knowledge of the particular hymns. The other (Finnish SOM) was trained with a set of Finnish folk songs and Lutheran hymns that are musically related to the spiritual folk hymns and would be familiar to the nonexperts. Of interest was whether the first network would weight veridical expectations more heavily than the latter. Thus, we have three different approaches to investigating the distinction between schematic and veridical expectation.

Schematic and veridical expectations are both top-down influences on melodic expectation because they both depend on prior experience. In the case of schematic expectation, it is the influence of other music with which the listener is familiar. In the case of veridical expectation, it is the influence of knowledge of the particular piece of music. In addition to top-down influences, bottom-up influences may play a role in melodic expectation. These bottom-up influences come from auditory perception and basic principles of perceptual organization. An example of the former is sensory consonance. An example of the latter is grouping by pitch proximity, timbral similarity, temporal contiguity, and so on (cf., Bregman, 1990). These bottom-up influences are assumed to come from the basic psychological system that processes sounds and organizes perceptual information into coherent groups and to function across music, speech, and environmental sounds.

Narmour's (1990) implication-realization model is based on such bottom-up principles. It is the most extensive and explicit model of melodic expectation to date. This model proposes a small number of perceptual (bottom-up) principles governing melodic expectation. These give rise to a system of melodic archetypes. The principles are based in part on Gestalt principles such as proximity, similarity, and good continuation. These principles are presumed to be universal and to operate independently of style knowledge. They have previously received empirical support (Krumhansl, 1991, 1995a, 1995b, 1997; Cuddy & Lunney, 1995; Thompson et al., 1997). It is therefore of interest to see whether the principles extend to another musical style, bearing in mind the possibility that the relative weighting of the underlying principles may vary across styles. In addition to testing the model against the behavioral data, we consider whether the neural network model abstracts similar principles. The network might abstract these principles because they may be reflected in the way music is structured; this has been shown for at least one musical style (Thompson & Stainton, 1998). We describe the implication-realization model and some recently proposed variants of it in some detail later.

## History and Background of Finnish Spiritual Folk Hymns

Before presenting the statistical, behavioral, and modeling approaches, we provide some background on the spiritual folk hymns and their cultural context. Although not all aspects of this description are directly relevant to the empirical research, we believe that such background is an important component of music cognition research in cross-cultural contexts. The corpus of music used in this study is drawn from the collection "Spiritual Songs for Devoted Souls" (Halullisten Sielujen Hengelliset Laulut, 1988, abbreviated here as HSHL). The style is the result of a collision between two different musical cultures, Lutheran hymns and Finnish folk songs. The majority of the melodies in the collection are variants of Lutheran hymns or Finnish folk songs. A few melodies were also adopted originally from German ballads. The origin of some melodies in the collection is still unknown, but it is generally believed that at least some of them were composed by anonymous individuals.

The HSHL collection has been used by many sects, such as "Herännäiset" (Pietists) and "Laestadiolaiset" (Laestadians), but it is generally associated with the Beseechers among whom the texts and melodies were created. Beseecherism is a religious sect that originated in the middle of the 18th century in southwest Finland (Figure 1). Typical of Beseecherism is religious conservatism, the tendency to preserve habits and traditions. The "Old Bible" and Carolinian liturgy is still in use by them as well as the "Old Hymn Book" from 1701. Conservatism is reflected also in their singing tradition. The singers try to preserve the songs as they were when they learned them in their childhood. This process is hindered by the fact that the songs are sung slightly differently from one village to another. The sect has only about 4000 members and they live in a narrow region in southwest Finland. Despite serious dogmatic conflicts that resulted in a division into two separate groups, Beseecherism is still a quite influential and vigorous religious force in Finland and forms a homogeneous cultural entity.

The first printing of HSHL was probably published around 1790 and contained only 26 hymn texts (Suojanen, 1984). Today the collection consists of 196 hymn texts. The HSHL collection has never included any kind of musical notation. The singers identify the melodies by reference to the texts in the "Old Hymn Book" or folk songs. Sometimes only an indication is found above the text that the hymn should be sung with its "own" melody.


**Fig. 1.** Region in Finland where Beseecherism is practiced. (Adapted from Suojanen, 1984, Appendix 1.)

The melodies have been learned orally from parents, grandparents, or from other people from the village. The main singing context of these spiritual folk hymns are “seurat” (religious meetings), funerals, weddings, and devotions at home.

The first book containing musical transcriptions of a few hymns was published in 1930 by Martti Merivirta (Merivirta, 1930). This book had little effect on the singing of Beseechers because few of them were able to read staff notation. The next books of musical notation were published in 1963 and 1988 (Halullisten Sielujen Hengelliset Laulut, 1963, 1988). The latter was based on recordings done by ethnomusicologist Erkki Ala-Könni (1922–1997) in 1956, and later by Jukka Louhivuori and Päivikki Suojanen (Louhivuori, 1988) during a 10-year period (1975–1985). The melodies used in the present study were taken from this last collection. The study of Finnish spiritual folk hymns started at the end of the last century by Ilmari Krohn (1867–1960), coinciding with the beginning of Finnish musicology (Krohn, 1899). The “first Finnish musicologist” developed his well-known method for classification of folk melodies in connection with these studies. The method was later adapted by the Hungarian musicologists and composers Béla Bartók and Zoltán Kodály (Kodály, 1959, p. 11).

From the point of view of the present study, the musical practice involving the spiritual folk hymns allows us to find a group of listeners who have had extensive experience singing the music from their childhood and a contrasting group with generally the same cultural background (and familiarity with Finnish folk songs and Lutheran hymns to which the spiritual folk hymns are related) but no familiarity with the particular hymns. The fact that spiritual hymn singing is based on an oral tradition is also beneficial. Because our expert group of hymn singers learned the melodies orally, we have good reason to believe that the observations made during the experiment are related to their ability to store the songs in memory. Thus, visual processing related to musical notation does not disrupt the observations. Another benefit is that the experimental observations can be compared with the extensive empirical data and musicological research done with spiritual folk hymns during this century.

As mentioned earlier, spiritual folk hymns are based mainly on the Lutheran hymn tradition and Finnish folk songs. Both traditions link spiritual folk hymns to European music culture. Beseechers varied the Lutheran hymns by adding grace notes and different kinds of auxiliary notes. The method is close to renaissance and baroque diminution techniques and jazz improvisation. After the addition of these notes, the Lutheran hymns sound very similar to traditional Finnish folk songs. Sometimes Lutheran hymns are so richly decorated that the original melody is difficult or impossible to recognize.

Typical of spiritual folk hymns is the existence of several more or less different variations. Even the same singer may sing different variations on different occasions. At the beginning of group singing, several variations can be heard at the same time. Toward the end of singing, the melody becomes more homogeneous, but normally a few variations will be heard even then. Also, a performance may shift between major and minor modes. For the present study, these variations are relevant because the expert subjects may have in mind several possible continuations. In fact, one of the expert hymn singers remarked that, although one particular hymn was familiar to her, she knew it in a slightly different form. Thus, the concept of "melody" is different from the typical Western concept. In spiritual folk hymns, a melody is not a fixed way of combining pitches and durations, but rather a set of possible ways to form a melodic whole.

The musical forms of spiritual folk songs are quite similar to other types of Finnish folk music. Most of the melodies are in binary or ternary form. In the present context, it is assumed that the form of the melody does not play a central role in melodic expectation, although this possibility cannot be tested without further studies. Spiritual folk hymns contain many melodic and rhythmic features with which almost every Finn is familiar. Although the songs are clearly part of Finnish music culture, they still contain remnants from musical styles that have disappeared from other types of Finnish music. A good example of this is the use of church modes, which have disappeared almost totally from other types of Finnish music. This fact has been used in the present study by choosing as test material some examples of modal spiritual folk hymns. Rhythmically, spiritual folk hymns are close to other Finnish folk songs, although some special types of songs—for example, melodies sung during prayer—are very complex rhythmically.

## Statistical Style Analysis

A number of studies suggest that statistical information is an important element in, for example, learning and perception. This has been found in music research and in other fields of study. In cross-cultural studies using North Indian (Castellano, Bharucha, & Krumhansl, 1984) and Balinese (Kessler, Hansen, & Shepard, 1984) music, the results showed that even inexperienced listeners were able to apprehend the basic tonal structure of the music by relying on the statistical distribution of the tones (see also Krumhansl, 1990). Similar results were obtained in a more recent study (Oram & Cuddy, 1995) that investigated how pitch-distributional information affects the perception of melodic sequences. It was found that both trained and untrained listeners were sensitive to the frequency of occurrence of tones. In particular, listeners appeared to use the distribution of tones in order to adapt to new tonal structures. The results also suggested that even in the presence of a familiar tonal organization, pitch-distributional information seems to be an important factor. Studies in other fields give further evidence for the importance of frequency information. For example, recent research suggests that language acquisition involves a significant amount of statistical learning (Saffran, Aslin, & Newport, 1996). The results indicate that infants are able to segment words from fluent speech by relying on the statistical cues found in the speech stream.

Clear statistical regularities have been found in note counts of real music. The data indicate that composers and improvisers emphasize the important tones of the tonality by using them more frequently, with longer duration, and more often in structurally important places than other tones (see Järvinen, 1995, 1997; Knopoff & Hutchinson, 1983; Krumhansl, 1990). It is also noteworthy that the distribution of the tones in, for example, tonal European art music and bebop-styled jazz conforms well with empirical results on the perception of tonality (Järvinen, 1995; Krumhansl, 1990). Given the important role of statistical information in listening, it seems that the communication between the listener and the person producing the music relies in part on the statistical properties of music. For experienced listeners, this information gives rise, for example, to more specific style expectations, but for inexperienced listeners, it gives cues about more general characteristics of the style and therefore it helps in learning that particular style (cf. Krumhansl, 1990; Oram & Cuddy, 1995).

Eighteen spiritual folk hymns from the collection HSHL were chosen for the statistical style analysis.[1] Half of the melodies were in major tonality and half were in minor tonality. (For this, the hymns in church modes were interpreted according to their relationship to minor or major tonality.) The selection of melodies was based on the following criteria. First, the melodies should be sung frequently and should be well known by experts in the style. Spiritual folk songs differ in the frequency of usage (Louhivuori, 1988). Second, the melodies should be representative examples of the style. The selected hymns represented a range of prototypical examples. Third, melodies that are especially complex rhythmically were not used as the present focus is on melodic expectation. Fourth, at least some of the melodies should contain features that are typical of spiritual folk hymns but not generally found in other types of Finnish folk music. As described below, the eight contexts chosen from the 18-hymn corpus for the behavioral experiment contained such distinctive features.

This study investigated the statistical distributions of tones, and two- and three-tone transitions in the corpus of 18 hymns.[2] All hymns were transposed to a common tonic of C. The actual experimental materials (which consisted of a subset of eight hymns) were also examined sepa-

1. The major hymns from HSHL were numbers 2a , 3b, 11a, 12a , 15a, 17, 20 , 32, and 54a. The minor hymns were numbers 10b, 24a, 31b, 33, 59b, 64a, 83a, 89b, and 192.

2. The data were analyzed with the Humdrum Toolkit (Huron, 1994) and some additional perl and shell scripts. These scripts can be found at http://www.jyu.fi/~tjarvine.

rately, but in the following discussion only the general results will be presented. In this study, we decided to use only the onsets of the tones instead of summing the durations of the tones, as has been done in some previous studies. The reason is that the performance of spiritual folk hymns does not always have a steady pulse or a clear global metrical structure, and therefore the listener might not be able to make a clear distinction between different note values. For example, in certain cases it can be impossible to decide whether a certain note is a quarter note with a fermata or a half note.[3] Consequently, no metrical analysis will be reported (cf. Järvinen, 1995). In any case, initial results showed that total durations and the number of onsets produced similar results.

## RESULTS

### Tone Distributions

Figure 2 shows the normalized statistical distribution of the tone onsets in both the major and minor hymns collapsed into one octave.[4] Both major and minor tone-frequency profiles show a clear preference for the tones in the diatonic scale; nondiatonic tones are almost nonexistent. In particular, the first, second, third, fourth, and fifth scale degrees are emphasized. It is noteworthy that in addition to the tones of the tonic chord (tones C, E♭/E, and G), the second scale degree (D) is also found quite frequently in these hymns. Similar emphasis on the second scale degree has been reported also in studies on other types of music (e.g., Järvinen, 1997; Knopoff & Hutchinson, 1983; Krumhansl, 1990). As an explanation, it has been suggested that this particular tone has an important melodic function as the upper neighbor tone of the root, and therefore it is used frequently as a melodic passing tone (e.g., Krumhansl, 1990, p. 69).

In general, these profiles accord well with both the statistical distribution of the tones obtained from other types of music and with psychological tests investigating the perception of tonality. Table 1 shows how the tone-frequency profiles correlate with tone profiles obtained in other studies (Järvinen, 1997; Knopoff & Hutchinson, 1983; Krumhansl & Kessler, 1982). In particular, the major key note counts acquired from tonal European art music appear to be nearly identical, $r(10) = .99$, with the statistical distribution of the tones in spiritual folk hymns. It is also noteworthy that

3. When investigating other types of music that have a continuous, steady pulse provided, for example, by an accompaniment (e.g., certain jazz or rock styles), it would seem to be reasonable to take into account the total duration of the tones (see, e.g., the note counts in Järvinen, 1995, 1997).

4. Onset frequencies and durations were similar; the correlations between the tone onset frequency profile with the profile that takes the tone durations into account was $r(10) = .98$ for minor hymns and $r(10) = .97$ for major hymns.

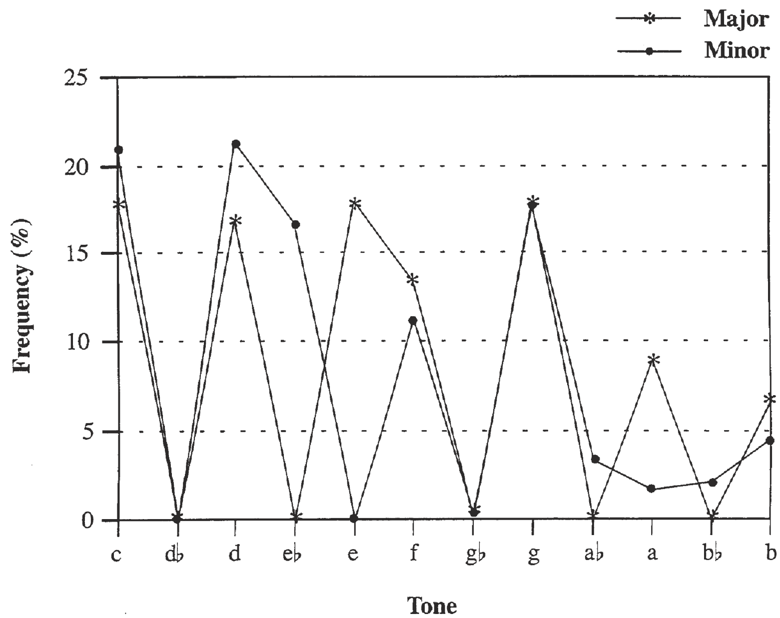


**Fig. 2.** Octave equivalent tone-frequency profiles for the spiritual folk hymns in major and minor collapsed into one octave (i.e., frequencies of pitch classes are shown).

such a different sounding style of music as bebop-styled jazz has such a high correlation, $r(10) = .96$, with the hymns. The correlations for the minor hymns are not quite as high, but a clear similarity still is evident.

The distribution of tones was also investigated without making the data octave equivalent. Figure 3 illustrates the frequencies of the tones in different octaves for both major and minor hymns. We use the convention that tones in the octave below middle C are denoted by capital letters, the tones in the octave containing middle C are denoted by lower-case letters, and

Table 1

**Correlations Between the Tone-Frequency Profiles Obtained from the Spiritual Folk Hymns and Previous Studies**

| Major | | Minor | |
|---|---|---|---|
| Study | Correlation | Study | Correlation |
| Krumhansl & Kessler | 0.88 | Krumhansl & Kessler | 0.79 |
| Knopoff & Hutchinson | 0.99 | Knopoff & Hutchinson | 0.89 |
| Järvinen | 0.96 | | |

The note counts come from studies by Knopoff and Hutchinson (1983) and Järvinen (1997), and the probe-tone rating profile comes from Krumhansl and Kessler (1982).

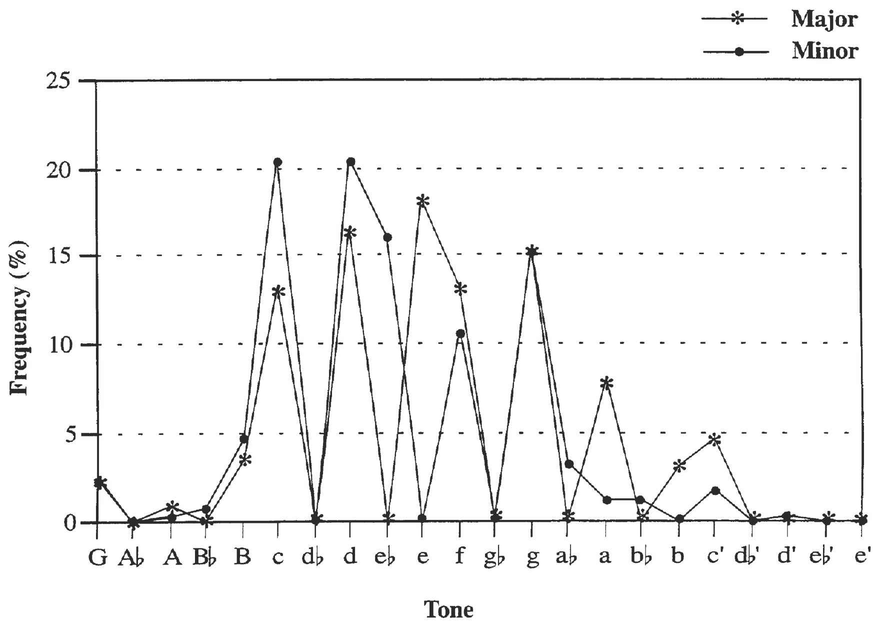


**Fig. 3.** Tone-frequency profiles for the spiritual folk hymns in major and minor with all octaves shown separately. Capital letters denote tones in the octave below middle C. Lower case letters denote tones in the octave beginning with middle C. Letters with primes indicate tones in the octave beginning with high C.

the tones in the higher octave are denoted by primes. The figure clearly shows that most of the melodic activity is in the middle octave. In addition, these data offer some clarification on the use of certain tones. For example, in the minor profile, the tone B is only used in the lower octave. This may be an indication of the further emphasis of c by using leading tone motion from B to c. In the major key profile, the relatively infrequent use of c is notable, but the reason for this may be that the high c' is also used as a reference point.

## Tone Transitions

In order to examine melodic aspects of the hymns, two- and three-tone transitions were examined. Tables 2 and 3 show, for the major and minor hymns, respectively, the two-tone transitions that appeared and the frequency with which they occurred; the three-tone transitions are listed in Appendix A. The data show that there are few wide leaps. Most of the transitions are stepwise movements or small intervals. The transition matrices also help to understand the tone-frequency profiles further. In particular, these transitions give some support for the explanation suggested

TABLE 2
**Two-Tone Transitions for the Major Spiritual Folk Hymns**

| Tone 1 \ Tone 2 | G | G♯ | A | A♯ | B | c | c♯ | d | d♯ | e | f | f♯ | g | g♯ | a | a♯ | b | c' | c♯' | d' | d♯' | e' |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| e' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| d♯' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| d' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | 1 | · | 1 | · | · |
| c♯' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| c'* | · | · | · | · | · | · | · | · | · | 1 | · | · | · | · | 2 | · | 13 | 8 | · | 1 | · | · |
| b | · | · | · | · | · | . | · | . | · | · | . | · | . | · | 13 | · | . | 4 | · | · | · | · |
| a♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| a | · | · | · | · | · | · | · | 5 | · | · | 5 | · | 22 | · | 5 | · | 4 | 2 | · | · | · | · |
| g♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| g | · | · | · | · | · | 1 | · | 2 | · | 7 | 27 | · | 18 | · | 15 | · | · | 9 | · | · | · | · |
| f♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | 1 | · | · | · | · | · | · | · |
| f | · | · | · | · | · | · | · | 4 | · | 42 | 8 | · | 13 | · | 4 | · | · | · | · | · | · | · |
| e* | · | · | · | · | · | 9 | · | 38 | · | 6 | 27 | · | 15 | · | 3 | · | · | · | · | · | · | · |
| d♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| d* | 5 | · | · | · | 6 | 25 | · | 16 | · | 22 | 2 | 1 | 11 | · | · | · | · | · | · | · | · | · |
| c♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| c | 1 | · | · | · | 9 | 13 | · | 17 | · | 16 | 1 | · | 1 | · | · | · | · | 1 | · | · | · | · |
| B | · | · | 4 | · | · | 8 | · | 7 | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| A♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| A | 3 | · | 1 | · | 1 | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| G♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| G | 1 | · | · | · | 2 | 8 | · | · | · | · | · | · | 1 | · | · | · | · | · | · | · | · | · |

The transitions that were used in the analysis of the behavioral data are marked by an asterisk (*).

for the relatively high frequency of the second scale degree (d), because in both the major and minor hymns, the melodic movement from d to c is one of the most frequent transitions. Furthermore, the minor matrix shows that the tone B functions in most cases as a leading tone to c. The major matrix, however, does not give similar evidence for the significance of c', for b does not seem to be used primarily as a leading tone to c'. However, the strong cadential transitions from g to c' give some support for the importance of c'. In connection with the behavioral data, presented next, only those transitions that match the experimental contexts were used; the transitions are marked with asterisks in the tables.

## Behavioral Experiment

Tone-to-tone expectation has been examined in detail in a number of previous behavioral studies. Some studies (Carlsen, 1981; Carlsen et al., 1970; Schmuckler, 1990; Thompson et al., 1997; Unyk & Carlsen, 1987) use production measures where the subject is presented a context and con-

Table 3
**Two-Tone Transitions for the Minor Spiritual Folk Hymns**

| Tone 1 \ Tone 2 | G | G♯ | A | A♯ | B | c | c♯ | d | d♯ | e | f | f♯ | g | g♯ | a | a♯ | b | c' | c♯' | d' | d♯' | e' |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| e' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| d♯' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| d' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| c♯' | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| c' | · | · | · | · | · | · | · | · | · | · | · | · | 4 | 4 | · | 1 | · | 1 | · | · | · | · |
| b | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| a♯ | · | · | · | · | · | · | · | · | · | · | · | · | 1 | 1 | 1 | · | · | 4 | · | · | · | · |
| a | · | · | · | · | · | · | · | · | · | · | · | · | 6 | · | · | 1 | · | · | · | · | · | · |
| g♯ | · | · | · | · | · | · | · | · | · | · | · | · | 14 | · | · | 4 | · | · | · | · | · | · |
| g* | · | · | · | · | · | 5 | · | 5 | 6 | · | 31 | 1 | 20 | 4 | 4 | 1 | · | 5 | · | · | · | · |
| f♯ | · | · | · | · | · | · | · | · | · | · | · | · | 1 | · | · | · | · | · | · | · | · | · |
| f | · | · | · | · | · | · | · | 6 | 28 | · | 4 | · | 13 | 6 | 2 | · | · | · | · | · | · | · |
| e | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| d♯ | · | · | · | · | · | 6 | · | 60 | 7 | · | 15 | · | · | · | · | · | · | · | · | · | · | · |
| d* | 3 | · | · | · | 7 | 34 | · | 8 | 38 | · | 9 | · | 10 | · | · | · | · | · | · | · | · | · |
| c♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| c* | 1 | · | · | 4 | 13 | 19 | . | 31 | 8 | · | · | · | 6 | 3 | · | · | · | · | · | · | · | · |
| B | 2 | · | · | · | 3 | 18 | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| A♯ | · | · | 2 | · | · | 2 | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| A | 2 | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| G♯ | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · | · |
| G | 2 | · | · | · | 1 | 8 | · | · | · | · | · | · | 2 | · | · | · | · | · | · | · | · | · |

The transitions that were used in the analysis of the behavioral data are marked by an asterisk (*).

tinues it by singing or playing an instrument. Other studies (Cuddy & Lunney, 1995; Krumhansl, 1991, 1995a, 1995b, 1997; Schmuckler, 1989) use perceptual measures where the subject is presented a context followed by a continuation tone that is judged (on a numerical scale) as to how well it fits with their expectations for continuation. These measures show general agreement (see especially Thompson et al., 1997). Perceptual measures have the advantage that they do not require singing or playing skills and are used in the present study. Studies on melodic expectation also differ in terms of the nature of the context used in the experiment. Some studies (Carlsen, 1981; Carlsen et al., 1970; Cuddy & Lunney, 1995; Krumhansl, 1995a; Thompson et al., 1997; Unyk & Carlsen, 1987) use simple two-tone contexts. Other studies (Krumhansl, 1991, 1995b, 1997; Schmuckler, 1989, 1990) use longer context sequences, usually extracted from musical compositions. Even the two-tone contexts elicit strong and regular expectation effects. For the purpose of examining the effects of style expertise, the present study used longer contexts taken from the hymn corpus. The results will, however, be compared with results from previous studies by using simple two-tone contexts.

The results will also be tested to see how well they conform to predictions of the implication-realization model (Narmour, 1990). The tests will consider the model as originally formulated and numerically coded (Krumhansl, 1991, 1995b), as modified in response to subsequent data (Krumhansl, 1995a), and as simplified through reanalysis of previous data (Schellenberg, 1996, 1997). These articles and Appendix B should be consulted for more technical aspects of the model. Briefly, Narmour proposed five underlying principles that generate a system of melodic archetypes. These principles apply to points in music when expectation for continuation is strong, that is, where the melody sounds incomplete. The last two tones at such points form what is called the *implicative interval*; the last tone of the implicative interval and the tone that follows is called the *realized interval*. The five principles describe perceptual principles that govern how the melody is expected to continue, that is, what realized interval is expected given the implicative interval. The expectations depend on the size of the implicative interval and its direction.

Two principles are governed by proximity; they state that a tone will be expected if it is close to the second tone of the implicative interval (Proximity) or the first tone of the implicative interval (Registral Return). The third principle (Registral Direction) is governed by good continuation; it states that a tone will be expected if the implicative interval is a small interval and the realized interval continues in the direction, or if the implicative interval is large and the realized interval reverses the direction. The fourth principle (Closure) states that a sense of closure is induced when the melody changes direction and when a large interval is followed by a smaller interval. The fifth principle (Intervallic Difference) is governed by similarity; it states that if the implicative interval is small a similarly sized realized interval is expected, whereas if the implicative interval is large, a smaller realized interval is expected.

These principles are precisely specified in the implication-realization model in terms of interval size. For example, the principle of Proximity states that the strength of expectation decreases with interval size down to 6 semitones (a tritone) and then remains at a constant low level. Registral Return is satisfied when the first tone of the implicative interval and the last tone of the realized interval are less than 3 semitones (a major second or smaller). This interval precision permits quantitative predictions that can be tested against behavioral data (Krumhansl, 1991, 1995b). The numerical coding is described in Appendix B. The quantified model was applied to melodic judgments made for contexts drawn from English folk songs, Chinese folk songs, and atonal songs (Krumhansl, 1995b), finding statistical support for most of the principles. Evidence for the principles also comes from studies by Cuddy and Lunney (1995), Thompson et al. (1997), and a statistical style analysis by Thompson and Stainton (1998) of

Bohemian folk-song melodies. Krumhansl (1997) found evidence that the principles also operate for nonadjacent tones on a higher hierarchical level, in addition to the tone-to-tone level.

Subsequent research has considered both additions to the model and simplifications of the model as originally quantified. One study (Krumhansl, 1995a) experimentally tested a comprehensive set of two-tone contexts (ranging from descending major seventh to ascending major seventh) followed by continuation tones over a two-octave range. A better fit to these data was found by a modified model that coded Proximity and Registral Return as linearly decreasing functions of interval size (without cut-offs at 6 and 3 semitones, respectively). These data also suggested that the model needed to be augmented by including effects of consonance. Specifically, a variable coding consonance was constructed for the realized interval and the interval formed by the first tone of the implicative interval and the second tone of the realized interval. It used empirical and theoretical values of consonance summarized in Krumhansl (1990, p. 57). This study also suggested the addition of variables coding when the realized interval is either a unison or an octave. The negative weight found for unisons indicates that repeated tones are not expected and the positive weight for octaves means that octaves are more expected than the other principles would predict. This revised model provided a better fit to the earlier data using contexts from English folk songs, Chinese folk songs, and atonal songs, and also the two-tone context data of Cuddy and Lunney (1995).

Schellenberg (1996) reanalyzed the data for the English folk songs, Chinese folk songs, and atonal songs, leading to a revised and simplified model. He proposed coding Proximity as a linear function (as had Krumhansl, 1995a) and a new coding of Registral Direction that assumed the principle only applies to intervals a perfect fifth or larger. This coding of Registral Direction is shown in Appendix B. Registral Return was coded as originally (Krumhansl, 1995b). He also omitted Intervallic Difference and Closure from the model. This simplified model provided an improved fit to the data of the three styles of music over the original formulation (Krumhansl, 1995b), but a worse fit for the Chinese folk songs and atonal songs compared to Krumhansl (1995a). In addition, the original coding provided a slightly better fit to the production data of Carlsen (1981) and Unyk and Carlsen (1987) than Schellenberg's (1996) revised model.

In a subsequent reanalysis of the same melodic continuation data, Schellenberg (1997) proposed a further modification. It combined the principles of Registral Return and Registral Direction into a single principle, called Pitch Reversal, shown in Appendix B. This, together with the linearly coded Proximity principle, constitutes his two-factor model. It again performed better than the original formulation (Krumhansl, 1995b) on all three styles of music, but worse on the Chinese folk songs and atonal songs compared with Krumhansl (1995a). These comparisons suggest that Schellenberg's (1996, 1997) modifications apply more to Western tonal music than other musical styles. More generally, reanalysis of these same data carries the risk of modeling specific details of the results that do not extend more generally. Consequently, the data from the present experiment will be used to test the various proposed extensions and simplifications of the implication-realization model.

## METHOD

### Subjects

There were two groups of subjects: experts and nonexperts. All 12 expert subjects belonged to the religious prayer movement Beseechers and were members of "The Youth Choir of the Beseechers" (*Rukoilevaisnuorten kuoro*). The experiment was conducted in the parish hall of Eura, a rural village, which is in the southwest part of Finland, before the choir had their normal rehearsal. The choir was paid a small compensation. The expert subjects' age range was from 14 to 41 years (average, 21.8 years). They had sung in the choir on average for 5.8 years and were active members of the choir (83% always participating in the rehearsals, 8% sometimes, 8% seldom) and they had active religious backgrounds (previously participating in religious meetings: 58% always, 42% often; and currently participating in religious meetings: 75% always, 8% often, 17% occasionally). They had heard and learned the spiritual folk hymns at an early age (on average beginning at 6.3 years old) and were taught by (in order of importance) the older people of the sect, mother, father, and/or grandparents. Of the expert subjects, 58% had other musical activities but only two had any formal music training (5 and 3 years). The familiarity with the hymns in the experiment was measured on a scale of 1 to 7. The experts were highly familiar with the style. They reported knowing well most of the hymns in the experiment (the average rating of familiarity with the individual hymns was 5.9).

Fourteen nonexpert subjects were undergraduate musicology majors from the Department of Music, University of Jyväskylä. Their average age was 22 years (range, 19–24 years), and they had participated in musical activities for an average of 17 years (had studied on average for 3.5 years professionally, 5.6 years music theory, 13.4 years on their major instrument). The experiment was a part of their music course. They were not familiar with the particular hymns; none of them reported recognizing any of the melodies (which would correspond to an average of 1.0 on the 7-point scale). However, on a question asking about their general familiarity with the style in question, they gave an average rating of 4.9. This is understandable because the hymns resemble some aspects of Finnish Lutheran hymns and folk songs with which these subjects are familiar. The music listening habits and musical tastes of the two groups were similar, although the choir members listened to more religious music than the nonexperts.

### Apparatus

The music was recorded on tape by playing the block of trials with Encore software on a Power Macintosh 6400/200 computer connected to an Ensoniq EPS 16 Plus keyboard synthesizer. The sound was set to an English horn sample (sampling rate, 44.1 kHz; 16-bit resolution). The experimental environment was a music classroom (for the nonexperts) and community/parish hall for the experts. Two (Yamaha NS-10M) speakers were positioned in the front of the room, and the trials were played on a Yamaha KX-W332 cassette deck connected to a Yamaha RX-350 amplifier. The experts heard the trials from a high-quality portable cassette player (JVC RC-X740).

### Stimulus Materials

Eight test contexts, shown in Figure 4, were chosen from the group of 18 spiritual folk hymns used in the statistical style analysis. The 18 hymns were first transformed to C major or C minor and then performed expressively on keyboard and recorded in MIDI. The experimental materials, the eight contexts and the experimental trials, were then derived from the MIDI files. The experimental trials consisted of the context plus an additional continuation tone. The continuation tones were the 22 chromatic scale tones in the range from G (the G below middle C) to e' (the e an octave and a third above middle C). The continuation tones were produced by changing only the original next pitch in the hymn, therefore keep-


**Fig. 4.** Test excerpts (eight selected melodies from *Halullisten Sielujen Hengelliset Laulut*) used in the behavioral experiment and the probe positions indicated by "(PP)."

ing the rhythm and key velocity the same as the original next tone in the melody. Because the listeners were run as groups, the order of the continuation tones was the same for all subjects (although different across contexts). It should be noted that previous studies (e.g., Krumhansl & Kessler, 1982) find minimal carry-over effects in a similar probe-tone task. (The exact replications, with the same contexts and different orders of probe tones, produced highly correlated results.) The trials were reproduced with an English horn sample, which had an amplitude envelope close to a sung human voice. They were played at a tempo of 120 beats/minute.

The first example in Figure 4 shows the complete hymn, and the excerpt that was chosen from it. A trial began with the excerpt shown in parentheses, followed by a single continuation tone at the probe position (indicated by PP). The final tone notated in the figure is the tone that actually occurs next in the melody. The probe positions were based on musicological knowledge of the spiritual folk hymns and Western tonal-harmonic music. Table 4 lists specific observations about the excerpts leading to their choice. These observations led us to believe that the probe positions would distinguish between experts and nonexperts. They were places where Western schematic knowledge would lead to expectations inconsistent with the tone that actually occurs next in the melody.

### Procedure

Subjects were given the instructions (in Finnish) in written form, and the instructions were also read and explained to them. They were told that they would hear fragments of melodies that start at the beginning of a phrase and stop in the middle of a phrase. They were asked to judge how well the continuation tone fits with their expectations about what might follow in the melody. These ratings will be called fitness judgments. The rating was made on a scale from 1 ("extremely bad continuation") to 7 ("extremely good continua-

TABLE 4
**Observations Relevant to Choice of Probe Positions**

| HSLH | Tonality | Meter | Observations |
|---|---|---|---|
| 33 | Minor (Dorian) | 2/4 | Unusual Dorian mode marked by the appearance in measure 6 of the Dorian sixth (a natural). |
| 59b | Minor | 3/4 | The melody predominantly moves step by step, but in measure 6 the melody suddenly moves by an upward sixth (to a♭). |
| 83a | Minor | 2/4 | The melody tends to move continuously until measure 6 when the melody stops moving; the same pitch (g) is repeated four times. |
| 89b | Minor (Dorian) | Alla breve | The beginning of the melody refers to C major, but in measure 3 the mode shifts to Dorian (with the appearance of e♭). |
| 2a | Major | 2/4 | Measure 6 contains a rare downward sixth (c'-e). The e is outside the previously established range. |
| 54a | Major | 2/4 | Measure 6 contains a downward jump to tonic (c), whereas the more typical scale step to d would be expected. |
| 12a | Major | 4/4 | Measure 13 contains a modulation to the dominant key with the appearance of f♯. |
| 20 | Major | 2/4 | In measure 6 an upward fourth (e-a) occurs rather than the more expected interval of a third to the dominant (e-g) outlining the tonic triad. |

tion"). Subjects were told that each melodic fragment may still sound incomplete even with the added tone, and they should not rate how well the continuation tones *completed* the fragments, but rather how well the continuation tones *continued* the fragments.

At the beginning of the experimental session, the listeners heard a practice block of five trials so they could ask any questions they might have about the instructions. Then the actual experiment began with a block of trials for each melodic fragment. At the beginning of each block of trials, the context (without any continuation tone) was presented three times so that listeners could become familiar with it. The experts were asked to rate how familiar they were with it prior to the experiment (also on a scale of 1 to 7). The nonexperts were asked following the experiment whether they recognized any of the contexts. A 10-s break was then followed by 24 trials. The first two trials were for practice and had the same continuation tones as the last two trials (trials 23 and 24); these were followed by the 22 experimental trials corresponding to each of the possible continuation tones. In each block of trials, the order of the continuation tones was random; blocks alternated between major and minor hymns. The experiment lasted about 65 min including the instructions and the practice trials. The two groups of subjects were tested separately as a group. Subjects marked their responses on answer sheets. At the end of the experiment the listeners filled out a questionnaire about their musical backgrounds.

## RESULTS

### Intersubject Correlations

The fitness judgments of the continuation tones across subjects were consistent across subjects within groups. All 132 intersubject correlations of the experts were highly significant, average $r(174) = .63$, $p < .0001$. (One block of trials from a subject was omitted because she apparently left out a response). Also, all of the 182 intersubject correlations of the nonexperts were highly significant, average $r(174) = .68$, $p < .0001$. The average intersubject correlation between the groups, $r(174) = .52$, was lower than the within-group correlations, showing greater consistency within than between groups. Because of the highly significant within-group correlations, the average results are used in most of the following analyses. The correlation between the average of the groups was $r(174) = .79$, $p < .0001$, showing a fairly high degree of agreement between groups despite their different knowledge of their hymns.

### Sample Results

Figure 5 shows a representative set of results for one of the Hymns, HSHL33, for the nonexperts and for the experts. As can be seen, the experts rated the next tone that follows the context, a, higher than the nonexperts rated it, possibly suppressing the rating for a♭ that would normally occur in the minor key. In contrast, the nonexperts rated a♭ (which is in the key of C minor) higher. The two subject groups clearly differ in their degree of veridical expectation. However, even though the experts knew the hymn and the correct continuation of the context, they rated some other tones rather highly. The following analyses are directed at discover-

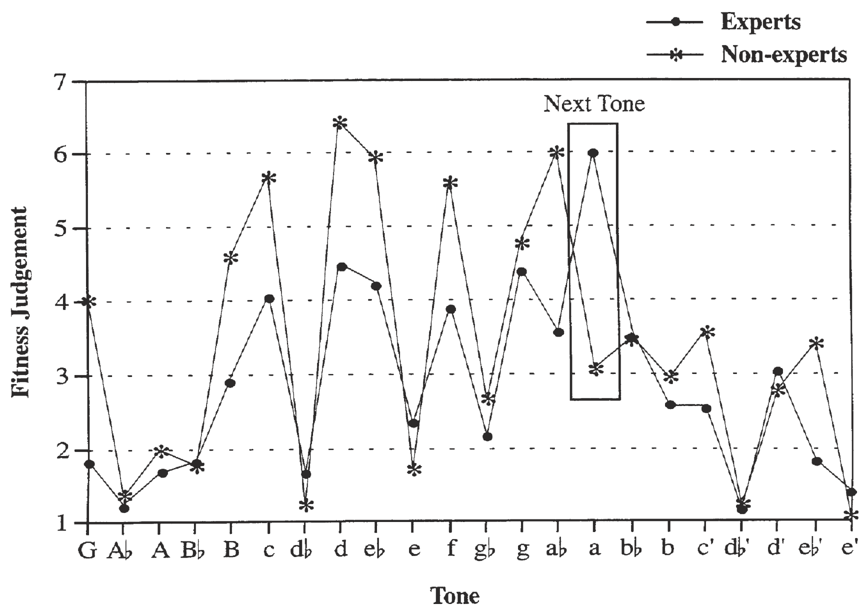


**Fig. 5.** Judgments of expert and nonexpert listeners in the behavioral experiment for one representative hymn, *Halullisten Sielujen Hengelliset Laulut* 33. The correct tone that occurs in the probe position is a, the Dorian sixth. Experts gave this tone a higher rating than nonexperts gave it. In contrast, the nonexperts gave a higher rating to the a$^{b}$ (which is in C minor).

ing the factors that influenced both sets of data, beginning with bottom-up principles and then turning to top-down expectations.

### Bottom-Up Principles: The Implication-Realization Model and Consonance

Four models, shown in Table 5, were tested against the data for each group, and also against the data for each subject. Model 1 coded the five principles of the implication-realization model as was done by Krumhansl (1995b). Model 2 coded the five principles of the implication-realization model as was done by Krumhansl (1995a), with the linear codings of Proximity and Registral Return and the two consonance variables (the consonance of the last tone of the context and the continuation tone, and the consonance of the second to last tone of the context and the continuation tone). Model 3 coded three principles of the implication-realization model as was done in the revised model by Schellenberg (1996). Model 4 is the two-factor version of the implication-realization model proposed by Schellenberg (1997).

Table 5 shows the multiple correlations for the four models, all of which were highly significant, and the adjusted $R^2$ values. These multiple correla-

TABLE 5
**Bottom-Up Principles: The Implication-Realization Model and Consonance**

| | Experts | Nonexperts |
|---|---|---|
| Model 1: Five implication-realization principles coded as in Krumhansl (1995b) | | |
| Multiple Regression | $R(5,170) = .53$*** | $R(5,170) = .43$*** |
| Adjusted $R^2$ | .26 | .16 |
| Registral direction | | |
| Intervallic difference | | |
| Registral return | | |
| Proximity | *** | *** |
| Closure | | |
| Model 2: Five implication-realization principles coded as in Krumhansl (1995a) and consonance | | |
| Multiple regression | $R(7,168) = .66$*** | $R(7,168) = .60$*** |
| Adjusted $R^2$ | .42 | .33 |
| Registral direction | | |
| Intervallic difference | ** | * |
| Registral return (linear) | | * |
| Proximity (linear) | *** | * |
| Closure | ** | |
| Consonance (with last tone) | *** | *** |
| Consonance (with second to last tone) | ** | *** |
| Model 3: Three implication-realization principles coded as in Schellenberg (1996) | | |
| Multiple regression | $R(3,172) = .54$*** | $R(3,172) = .45$*** |
| Adjusted $R^2$ | .28 | .19 |
| Registral direction (revised) | | |
| Registral return | | |
| Proximity (linear) | *** | *** |
| Model 4: Two implication-realization principles coded as in Schellenberg (1997) | | |
| Multiple regression | $R(2,173) = .53$*** | $R(2,173) = .44$*** |
| Adjusted $R^2$ | .27 | .18 |
| Pitch reversal | | |
| Proximity (linear) | *** | *** |

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

tions were computed for the group averages. Model 2 outperformed the others, which performed about equally to one another. All four models were also applied to each individual subject's data. To test for differences between models, the individual subject's multiple correlations were entered into an analysis of variance with the result, $F(3,75) = 170.8$, $p < .0001$, showing highly significant differences between the models. This difference was that Model 2 provided a better fit than the other three models, which were not significantly different from one another.

Table 5 also shows which principles made significant contributions to the fit of the models. For Models 1, 3, and 4, only Proximity made a significant contribution (in either the original or linear forms). Critically, neither the revised Registral Direction variable (Schellenberg, 1996) nor the proposed Pitch Reversal variable (Schellenberg, 1997) was significant. In Model 2, Proximity was again strong, and some support was also found for Intervallic Difference, Registral Return, and Closure. In addition, both consonance variables made significant contributions. In what follows, only Model 2 is considered.

Figure 6 shows the simple correlations between the seven variables of Model 2 and the data of the experts and nonexperts. The graph, in which the dashed line indicates the correlation that is significant at $p < .05$, generally confirms the results of the multiple regression. In the simple correlations, Intervallic Difference, Registral Return (linear), and Proximity (linear) were relatively strong and significant. The two consonance variables were weakly significant. Registral Direction and Closure were not significant. It is important to note that the results for both groups of subjects were quite similar in the correlations with the bottom-up variables. That is, the relative strengths of the principles did not strongly depend on familiarity with the music.

## Top-Down Influences: Schematic to Veridical Expectations

The following variables were used to create a continuum between schematic and veridical expectations: the two-tone context data (in which listeners judged continuation tones following a two-tone interval, Krumhansl, 1995a), the major and minor key profiles (Krumhansl & Kessler, 1982), the distribution of tones in the corpus of 18 hymns, the distribution of two-tone transitions in the corpus, the distribution of three-tone transitions in the corpus, and the correct next tone in the hymn. The values for the tone

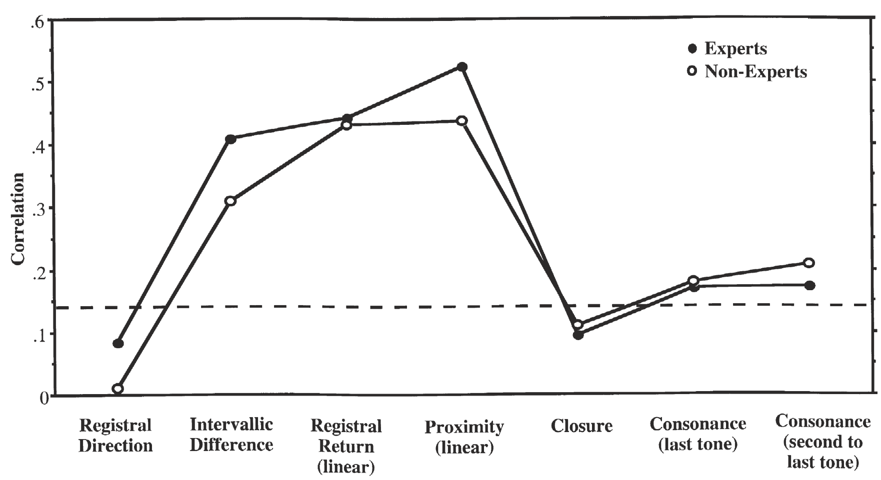


**Fig. 6.** Simple correlations between the bottom-up principles and the judgments of experts and nonexperts. The dashed line indicates the correlation with $p < .05$.

distributions were the normalized frequencies shown in Figure 3. The values for the two-tone transitions were the conditional probabilities of each continuation tone given the last tone of the context (derived from the data in Tables 2 and 3). The values for the three-tone transitions were the conditional probabilities of each continuation tone given the last two tones of the context (derived from the data in Appendix A). The correlations were computed between each of the variables and the data of each of the groups.

The following analysis focused on the differences between the groups in terms of the variables on the schematic to veridical continuum. Figure 7 shows the differences between the correlations for the groups. Negative values mean that the correlation is stronger for the nonexperts, and positive values mean that the correlation is stronger for the experts. (The values of the correlations are shown in Table 7, discussed later.) As can be seen, the nonexperts showed stronger effects of the more schematic variables, whereas the experts showed stronger effects of the more veridical variables. Individual subject correlations were compared to find which variables were significantly different between groups. The correlation with the two-tone context data (Krumhansl, 1995a) was stronger for the nonexperts, $r(174) = .71$, than the experts, $r(174) = .63$, and this difference was statistically significant, $F(1,24) = 14.33$, $p < .0009$. Similarly, the correlation with the major and minor profiles were stronger for the nonexperts, $r(174) = .51$, than the experts, $r(174) = .44$, and this difference was statistically significant, $F(1,24) = 6.31$, $p < .02$. In contrast, at the other end of the continuum, the correlation with the three-tone transitions was stronger for

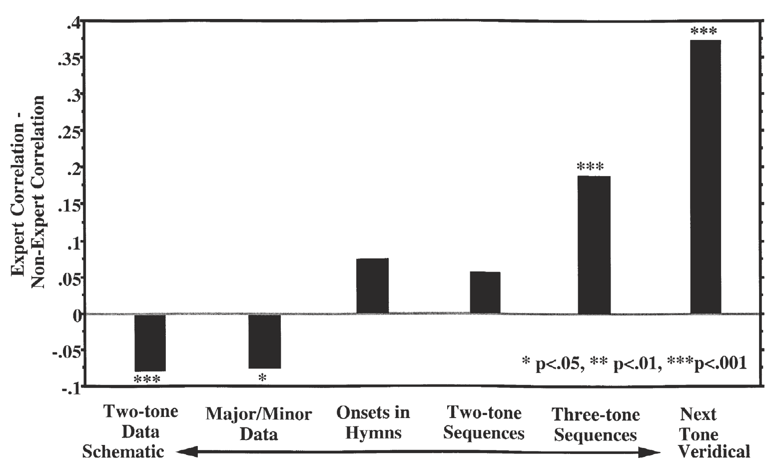


**Fig. 7.** Continuum of variables ranging from schematic to veridical expectancy. The figure shows the differences between the correlations for expert and nonexpert listeners. A negative value means the variable correlated more strongly with the nonexperts' judgments; a positive value means the variable correlated more strongly with the experts' judgments.

the experts, $r(174) = .59$, than the nonexperts, $r(174) = .41$, and the difference was statistically significant, $F(1,24) = 35.91$, $p < .0001$. Finally, the correct next tone were stronger for the experts, $r(174) = .52$, than the nonexperts, $r(174) = .15$, and the difference was statistically significant, $F(1,24) = 127.58$, $p < .0001$. Because the correct next tone variable is highly skewed, this last difference was substantiated by an analysis of the judgments given to the correct next tone. These were given higher ratings by the experts (average of 6.04) than the nonexperts (average of 4.62), $F(1,24) = 17.79$, $p = .0003$.

### Bottom-Up and Top-Down Influences

Other tests of the implication-realization model (Krumhansl, 1991, 1995a, 1995b; Schellenberg, 1996, 1997) have included a variable coding tonality (for Western music, the Krumhansl & Kessler, 1982, major and minor key profiles). For comparison with those results, a multiple correlation was conducted using the entire set of bottom-up and top-down principles (excluding the two-tone context data of Krumhansl, 1995a, because it did not cover all cases that appeared in the present experiment). Table 6 shows the results. The multiple correlations were highly significant for both groups, although the correlation was considerably higher for experts than nonexperts (when computed for individuals, $F(1,24) = 30.36$, $p < .0001$).

TABLE 6
**Bottom-Up Principles and Top-Down Influences**

| | Experts | Nonexperts |
|---|---|---|
| Model 2: Five implication-realization principles coded as in Krumhansl (1995a) and consonance with top-down influences | | |
| Multiple regression | $R(12,163) = .90$*** | $R(12,163) = .77$*** |
| Adjusted $R^2$ | .80 | .56 |
| Bottom-up principles | | |
| Registral direction | | |
| Intervallic difference | * | |
| Registral return (linear) | * | ** |
| Proximity (linear) | ** | |
| Closure | * | |
| Consonance (with last tone) | *** | |
| Consonance (with second to last tone) | | |
| Top-down expectations | | |
| Major/minor key profiles | *** | *** |
| Tone onsets | *** | ** |
| Two-tone transitions | | |
| Three-tone transitions | * | |
| Next tone | *** | |

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

This difference is explained in part by the contribution to the experts' judgments of the higher-order transitions and the correct next-tone variable. Both groups of subjects had significant contributions of both the major/minor key profiles and the onsets in the hymns, indicating sensitivity to both schematic Western tonal hierarchies and the distribution of tones in the hymn style.

## Self-Organizing Map Simulations

The subjects' ratings of the continuation tones were assumed to be based on a conceptual representation of structures typical of the musical style used. These representations have been suggested to develop through repeated exposure to music and the regularities inherent in it (Bharucha, 1987; Krumhansl, 1990; Oram & Cuddy, 1995). In the present study, the development of such conceptual representations was simulated with a neural network model of the self-organizing map type (Kohonen, 1997), denoted SOM. In previous work, Bharucha and Todd (1989) modeled the formation of schematic and veridical expectations of tonal structure with a sequential neural network based on the design by Jordan (1986). They showed that the network learned a novel tonal sequence more quickly if the sequence conformed to familiar patterns (specifically, frequent chord transitions) in previously learned sequences. Moreover, using the technique of cascading, they showed that the number of cascaded steps needed to reach asymptote was much smaller for expected transitions than unexpected ones.

The SOM is an artificial neural network that, through repeated exposure to a distribution of input vectors, is capable of developing a generalized representation of the distribution of input vectors. The SOM consists of a set of units that usually forms a planar array. Each unit is associated with a reference vector with the number of dimensions equal to that of the input vectors. When presented with a set of input vectors, the SOM settles into a configuration where the reference vectors optimally approximate the set of input vectors. Moreover, the SOM provides a topographical mapping of the set of input vectors. The mapping is topographical in the sense that similar input vectors are mapped to nearby units.

The process of self-organization can be shown to be equivalent to minimizing the average quantization error for the set of input vectors for the application (Kohonen, 1997, pp. 127–130), where quantization error is defined to be the distance between the input vector and the closest reference vector on the map. The result of minimizing the quantization error is shown schematically at the top of Figure 8. The input vectors used for training the SOM are indicated as Xs. The reference vectors of the trained SOM are shown by black circles. The reference vectors approximate the distribution of input vectors, minimizing the distances between the input vectors and their closest reference vectors.

After the SOM has been trained, the quantization error provides a measure of how well a given input fits with the generalized representation obtained from the training data. In the present study, the quantization error provides a way of measuring how well each sequence used in the behavioral experiment fits with the trained SOM. The bottom of Figure 8 shows schematically the quantization errors of three sequences. The input vectors are shown as Xs and the arrows point to the closest reference vectors. The quantization errors are the distances shown in the figure. It was assumed that input vectors representing sequences typical of the musical style learned by the map would have a small quantization error, and vice versa. Consequently, the goodness-of-fit of a given sequence to the SOM, called the fitness measure, was defined to be the negative of the quantization error of the input vector for the melodic sequence.

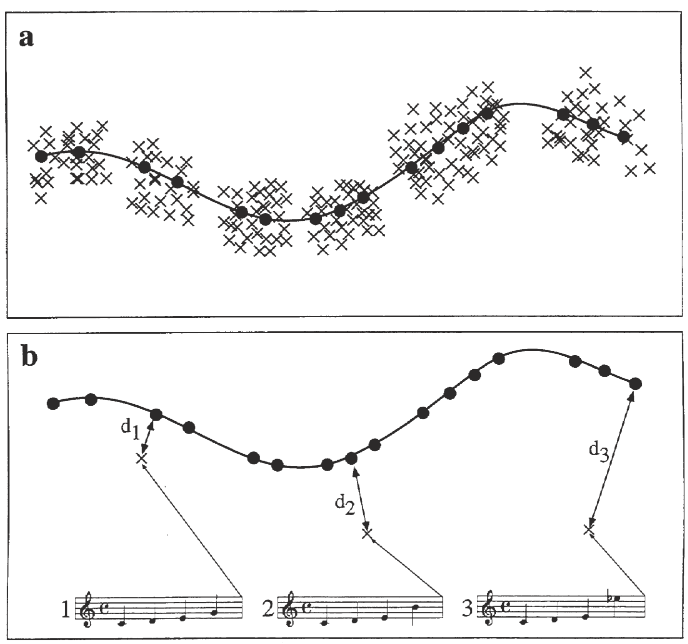


**Fig. 8.** A schematic, low-dimensional representation of a trained self-organizing map (SOM) and the quantization error. (a) The input vectors used for training the SOM are denoted by Xs, and the reference vectors of the SOM by black circles. The reference vectors approximate the distribution of the input vectors. (b) The quantization errors of three hypothetical sequences, 1, 2, and 3. The vector representations of the sequences, obtained from the preprocessing stage, are represented by Xs. The quantization error of sequence 1, $d_1$, is the lowest, whereas that of sequence 3, $d_3$, is the highest. Consequently, sequence 1 has the highest fitness value, and sequence 3 the lowest.

Kohonen (1997) stresses the importance of preprocessing the input. Figure 9 shows a schematic diagram of the preprocessing stages of the present model. The input data used in the simulations consisted of digitized recordings of musical performances. In order to use the data as input to the SOM, it was necessary to preprocess the sound files to extract significant features. The present focus was on modeling melody perception, so the most essential feature of the tones to be extracted was pitch. Because the melodies were monophonic, this could be carried out using a relatively simple algorithm based on periodicity analysis. A mathematical description of this algorithm is provided in Appendix C. The pitch extraction stage yielded as output a 22-dimensional vector (corresponding to the 22 tones from G to e') every 40 ms representing the instantaneous pitch and amplitude of the melody.

Like most neural networks, the basic SOM is inherently static in the sense that it is not capable of processing sequential input. Because the temporal order of tones is central to melody perception, this order should be represented in the model in some way. One way of dealing with this is to convert sequential information into static form by constructing a set of memory trace vectors and using these vectors for training the SOM. In the

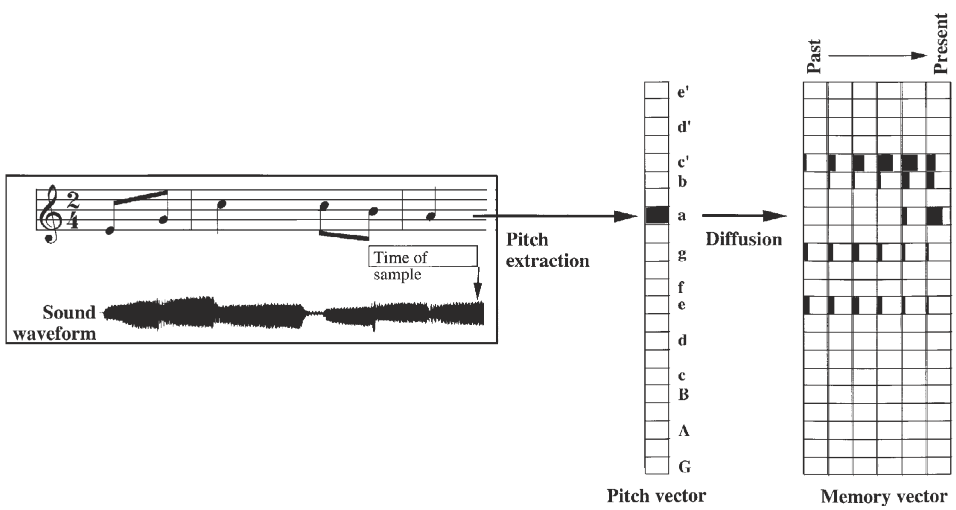


**Fig. 9.** Preprocessing of the input data. The pitch and memory vectors shown in the figure were obtained at the point of time indicated on the sound waveform image. At this point, pitch a is sounded. The component of the pitch vector representing this pitch has a high value, whereas all the other components have zero values. In the memory vector, the most recent pitches are represented in the rightmost column. This column contains a high value at pitch a and somewhat lower values at pitches c' and b. Correspondingly, the oldest events of the memory span are represented in the leftmost column. This column contains the highest value at pitch e and lower values at pitches g and c'.

present study, the memory trace vectors were constructed using a diffusion mechanism. Governed by equations similar to those of physical diffusion, this mechanism is based on spreading activation in a chain of units. A mathematical description of the diffusion memory algorithm is given in Appendix C. As seen in Figure 9, the memory trace vectors in the present model consisted of a chain of six units. The diffusion memory stage yielded as output a 132-dimensional vector (6 units each consisting of the values for 22 tones) every 40 ms representing the melodic progression during the last few seconds. In this mechanism, the temporal dimension is thus mapped onto a spatial dimension; the position of activation indicates when the respective events occurred.

In previous neural network models of music cognition, memory has mainly been modeled using leaky integrators (e.g., Bharucha & Todd, 1989; Gjerdingen, 1989, 1990; Griffith, 1994; Leman, 1994, 1995). The primary advantage of using diffusion memory instead of leaky integrators is that the temporal order of events can be represented with a finer resolution. For instance, in diffusion memory, recent, weak signals are represented differently from distant, strong signals, which does not necessarily hold true for leaky integrators. Furthermore, the temporal resolution of diffusion memory varies so that recent events are represented with a finer temporal resolution than distant ones. This property seems appealing, because it corresponds qualitatively with human memory.

Two different corpora of melodies were used for training the SOM.[5] The first corpus consisted of four Finnish folk songs and four Lutheran hymns, chosen to be representative of the styles.[6] Half of the hymns and folk songs were in major and half were in minor. The resulting SOM will be called the Finnish SOM, representing general knowledge of Finnish music in styles related to the spiritual folk hymns. The second corpus consisted of the eight contexts used in the behavioral experiment, including the correct continuation tones. It was assumed that the knowledge acquired by the Finnish SOM would emphasize more general features of the musical styles to which the hymns are related, whereas the knowledge acquired by the Hymn SOM would emphasize more specific features of the contexts used in the

5. The simulations were carried out with an SGI Power Onyx computer, located at the Center for Scientific Computing in Espoo, Finland.

6. The Lutheran hymns were *Enkeli taivaan* (Angel of heaven), Martin Luther, 1535 (hymn book number 21, as designated by the Finnish Evangelic Lutheran church in 1986), *Jumala ompi linnamme* (God is our fortress), Martin Luther, 1528 (hymn book number 170), *Lienenkö outo* (Am I a stranger?), German, 1675 (hymn book number 281), and *Vakaana Herran teitä* (Steady on the Lord's path), Johan Olof Wallin, 1816 (hymn book number 393). The Finnish folk songs were *Minä olen Härmän Kankaanpäästä* (I am from Kankaanpää, Härmä), *Ol' kaunis kesäilta* (It was a beautiful summer evening), *Aamulla varhain* (Early in the morning), and *Taivas on sininen ja valkoinen* (The sky is blue and white).

experiments. If so, the fitness measures produced by the Finnish SOM would correspond more to schematic expectations, whereas those produced by the Hymn SOM would correspond more to veridical expectations.

To train the SOM, the input sound file was first preprocessed using pitch extraction and diffusion memory. The effect of the memory length parameter (see Appendix C) was studied by constructing several sets of diffusion memory vectors with memory length parameter values ranging from 0.5 to 6.0 s and training the SOM with each set. Preliminary tests found that the memory length parameter had a significant effect on the final configuration of the SOM; the value of 1.5 s yielded the highest correlation between the responses of the SOM and the subjects. With this value of the memory length parameter, the last unit on the diffusion chain attains its maximal activation level 1.0 s after the presentation of an input. After that, it starts to decay, reaching the level of 50% of the maximum activation at 2.0 s and 10% of the maximum activation at 3.0 s after the input. This period agrees with what is known about the length of the auditory sensory memory or echoic store. Estimates of the length of this store range between about 2 and 5 seconds (e.g., Darwin, Turvey, & Crowder, 1972; Fraisse, 1982; Treisman, 1964). Further, the memory span used in the present simulations also agrees with temporal integration periods used in other simulations of similar type (e.g., Huron & Parncutt, 1993; Leman, 1995). The results reported below were obtained by using this value.

In each of the simulations, the SOM consisted of 30 by 30 units. The number of units was selected to be a compromise between the resolution of the network and the time needed for training. The Euclidean metric was used to measure the distance between vectors. The SOMs were trained in two phases. The first phase, the ordering phase, consisted of 10,000 cycles, where each cycle included choosing a vector randomly from the set of input vectors and adapting the SOM to this vector. The Euclidean metric was used to measure the distance between vectors. During the ordering phase, the learning rate (which governs how rapidly the unit vectors are adapted to the input vectors) was linearly decreased from 0.50 to 0.10. Similarly, the neighborhood radius was decreased linearly from 15 to 3, where the unit of measure is the distance between two adjacent units of the map. The second phase, the convergence phase, consisted of 500,000 cycles, during which the learning rate was linearly decreased from 0.10 to 0 and the neighborhood radius from 3 to 0. After the SOM was trained, its response to each of the stimulus sequences used in the behavioral experiment was measured. For this, the sound files containing the stimuli were preprocessed as described earlier. For each trial of the experiment, the vector used for testing the SOM was the diffusion memory vector taken at the end of the stimulus sequence. For each stimulus vector, the fitness measure of this vector was then determined as was shown schematically at the bottom of Figure 8.

RESULTS

## Intermodel Correlations

The correlation between the two SOM models was $r(174) = .38$. Although highly significant, $p < .0001$, this correlation is considerably lower than that between the subject groups, $r(174) = .79$. This can be understood because the subject groups had largely similar musical experiences, except for knowledge of the spiritual folk hymns, whereas the two SOMs were exposed to nonoverlapping corpora.

## Bottom-Up Principles: The Implication-Realization Model and Consonance

The fitness measures of the Hymn SOM were analyzed in terms of the implication-realization model and consonance (using Model 2 of Table 5). Recall that this model coded Proximity and Registral Return as linear distance from the last tone of the context and the second to last tone of the context, respectively. The multiple correlation was highly significant, $R(7,168) = .49$, $p < .0001$, adjusted $R^2 = .21$. Figure 10 shows the simple correlations between the seven variables of the second model and the fitness measures of the Hymn SOM. As with the behavioral data, Intervallic Difference, Registral Return (linear), and Proximity (linear) were relatively strong and significant. Again, Registral Direction and Closure were not significant. Here, however, only one of the consonance values was signifi-

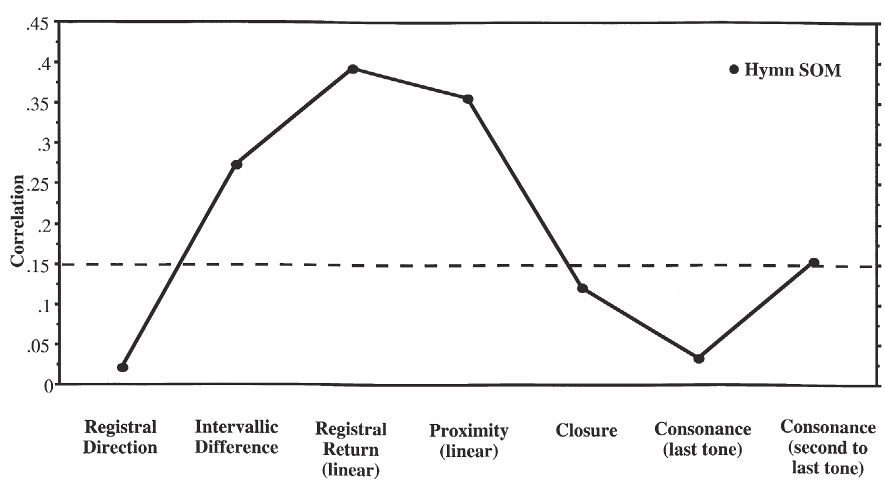


**Fig. 10.** Simple correlations between the bottom-up principles and the fitness measures of the Hymn self-organizing map (SOM). The dashed line indicates the correlation with $p < .05$. Comparison with Figure 6 shows that the Hymn SOM weighted the principles similarly to the expert and nonexpert listeners.

cant, and only weakly. In general, however, the SOM fitness judgments showed the same relative contributions of the bottom-up principles as the behavioral data of both groups (compare Figure 6).

### Top-Down Influences: Schematic to Veridical Expectations

This analysis focused on the differences between the Hymn SOM and the Finnish SOM in terms of the variables on the schematic to veridical continuum. The correlations were computed between each of the top-down variables and the two SOMs. Figure 11 shows the differences between the correlations for the SOMs. As can be seen, these generally parallel the differences found between the experts and the nonexperts (compare Figure 7). The Hymn SOM showed stronger effects of the more veridical variables than the Finnish SOM, especially the three-tone transitions and the next tone variable. These differences are comparable in magnitude to those found in the behavioral data. (The correlations, shown in Table 7, are discussed later.) However, unlike the behavioral data, the SOM differences were negligible for the more schematic variables. The correlations with the two-tone context data (Krumhansl, 1995a) were $r(174) = .39$ and .41 for the Finnish and Hymn SOMs, respectively (both $p < .0001$). The correlations with the major/minor data (Krumhansl & Kessler, 1982) were both $r(174) = .30$, ($p < .0001$). These two schematic variables are both experimental results for schematic Western contexts; it is not clear whether these would

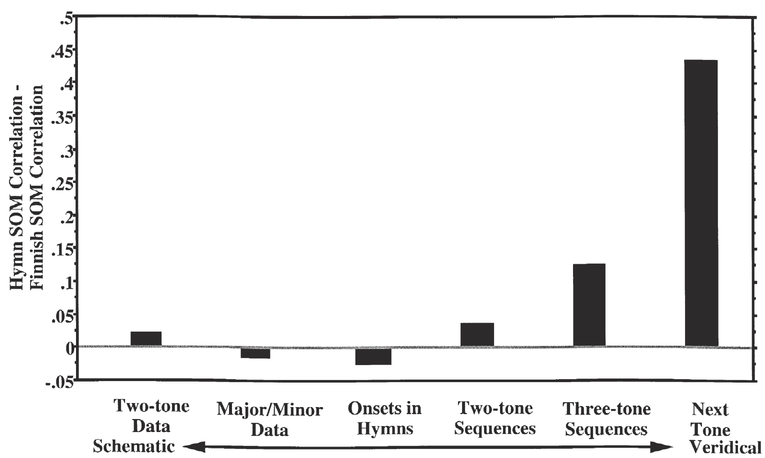


**Fig. 11.** Continuum of variables ranging from schematic to veridical expectancy. The figure shows the differences between the correlations for the Hymn self-organizing map (SOM) model and the Finnish SOM. A negative value means the variable correlated more strongly with the Finnish SOM; a positive value means the variable correlated more strongly with the Hymn SOM model. The similarity with Figure 7 shows the two SOM models characterize some of the differences between expert and nonexpert listeners.

TABLE 7
**Convergence of Approaches**

| | Behavioral Experiment | | Neural Network Model | |
|---|---|---|---|---|
| | Nonexperts | Experts | Finnish SOM | Hymn SOM |
| Statistical style analysis | | | | |
| Tone distributions | .64*** | .72*** | .49*** | .50*** |
| 2-Tone transitions | .50*** | .56*** | .40*** | .42*** |
| 3-Tone transitions | .41*** | .59*** | .32*** | .45*** |
| Correct tone | .15 | .52*** | .05 | .49*** |
| Behavioral experiment | | | | |
| Nonexperts | — | .79*** | .45*** | .48*** |
| Experts | .79*** | — | .43*** | .63*** |

*** $p < .0001$

be more similar to the Finnish folk songs and Lutheran hymns (of the Finnish SOM) or the Finnish spiritual folk hymns (of the Hymn SOM). As was seen in Table 1, the onsets of tones in the hymns were very similar to those of other collections of Western music. Thus, no substantial difference between the models for these variables would be expected, consistent with the results.

## Convergence Between the Three Approaches

Table 7 shows the correlations between the results of the three approaches. The top of the table shows the results of the statistical style analysis (going from tone distributions, to two- and three-tone transitions, to the correct next tone) compared with the behavioral experiment and the SOMs. As can be seen, most of the correlations were highly significant. Consider first the correlations for the nonexperts in the first column; these dropped off substantially for the higher order statistics and the correct next tone. These results are highly comparable with the correlations for the Finnish SOM shown in the third column. Again, the correlations decreased for the higher order statistics and the correct next tone. In contrast, the correlations for the experts in the second column and the Hymn SOM in the fourth column remained relatively high for all the variables. Thus, when compared with the results of the statistical style analysis, the experts and the Hymn SOM were similar, as were the nonexperts and the Finnish SOM.

Consider next the correlations between the behavioral data and the SOM models shown in the rest of the table. The fitness measure of the Hymn SOM correlated more strongly with the experts, $r(174) = .63$, $p < .0001$, than the nonexperts, $r(174) = .48$, $p < .0001$. When computed for individual subjects, the difference was significant, $F(1,24) = 40.19$, $p < .0001$. In contrast, the fitness measure of the Finnish SOM correlated about equally strongly with the experts, $r(174) = .43$, $p < .0001$, and the nonexperts, $r(174) = .45$, $p < .0001$. When computed for individual subjects, this difference was not significant, $F(1,24) = 1.42$, n.s. This lack of difference between groups for the Finnish SOM can be understood because both groups of listeners would be expected to be equally familiar in general with Finnish music (including Finnish folk songs and Lutheran hymns). Expertise was associated only with higher correlations with the Hymn SOM with no difference for the Finnish SOM (the interaction between SOM and group was $F(1,24) = 26.02$, $p < .0001$). Thus, the Hymn SOM appears to have abstracted some of the same information that the expert listeners used in making their continuation judgments.

## General Discussion

This study, using a special corpus of music, allowed us to investigate a number of issues from a variety of perspectives and with a variety of methods. It also suggested certain questions that would be interesting to investigate more systematically and in more detail in future research. The musical practice concerning Finnish spiritual folk hymns allowed us to obtain behavioral responses from listeners with generally similar cultural and musical backgrounds. However, only one group had extensive knowledge of the Finnish spiritual folk hymns, and they were familiar with the particular materials used in the experiment. This difference between groups produced clear and interpretable patterns in the melodic expectation judgments that could be traced to stylistic features uncovered through both ethnomusicological studies and statistical style analyses.

The melodic continuation ratings for the experts reflected more strongly veridical expectations, specifically three-tone transitions and the next tone that occurs in the hymn, than the nonexperts. The opposite was true for schematic expectations, such as those found in studies using two-tone contexts (Krumhansl, 1995a) and schematic key-defining contexts (Krumhansl & Kessler, 1982). Thus, the type of top-down knowledge distinguished between the two groups of listeners. In contrast, the bottom-up principles, as formulated in the implication-realization model (Narmour, 1990), had similar influences for the two groups of listeners. In other words, the relative influence of the principles on the behavioral data was largely independent of prior knowledge of the music. This suggests that musical expertise is associated with two specific psychological processes: sensitivity to higher order statistical properties of the music and ability to recognize fragments of pieces and generate correct (or veridical) expectations for continuation.

It should be noted, however, that the present study confounds two factors: style knowledge and familiarity with the particular hymns. It would be interesting, therefore, to separate these factors in future studies. This could be done by selecting some hymns unfamiliar to the style experts and others familiar to them. The present results showed two trends suggesting this would be a fruitful approach. Some of the hymns were somewhat less familiar to the experts than others. As would be expected, the intergroup correlations tended to be higher for the less familiar hymns and the veridical expectations were weaker. Another study using the same methods (Krumhansl, Toivanen, Eerola, Toiviainen, Järvinen, & Louhivuori, 1999) did separate style knowledge and familiarity. That study used as stimulus materials Sami yoiks from Lapland. That study had three groups of listeners: Sami, Finnish, and Western. Some of the yoiks were unfamiliar to the Sami listeners, whereas some of the other yoiks were familiar to the Finnish listeners. None of the yoiks was familiar to the Western listeners. The results suggested that knowledge of the particular musical materials can compensate for the lack of style knowledge, at least as measured by melodic continuation judgments. At a more general level, this kind of approach promises insights into the kinds of musical experience that lead to musical expertise.

The behavioral responses also showed interesting patterns that may relate to the history of spiritual folk hymns and their stylistic features. The use of modes is a typical feature of the hymns but is relatively rare in most other kinds of music. Some of our materials used unusual modes, resulting in distinct patterns for the experts and nonexperts. A more systematic study of effects of modes would be desirable and, in addition, may be interesting from a historical point of view. The statistical style analysis in the present study showed, for example, that the tone B occurs relatively infrequently in the corpus. This may be traced back to the hexachordal system (C, D, E, F, G, A) in earlier musical practice. Another stylistic feature that could be examined in greater detail is the existence of many variations of spiritual folk hymns. One consequence of this is that expectations may be more variable, with a greater number of tones accepted as good continuations, at points in the hymns where many variations exist. One effect of variations found here is suggestive. Numerous hymns exist in both major and minor modes and, indeed, a performance may shift between the two modes. Consistent with this, the responses of the experts in this study less sharply distinguished between major and minor modes than did the nonexperts.

The present study also suggests the desirability of relying more heavily on statistical style analysis in future studies, both for selecting the musical materials and for analyzing the results. The purpose of statistical analysis of the spiritual folk hymns in this study was to get precise information about the underlying regularities in this style. This findings were then used to analyze the behavioral data and the SOM model in order to find out if they were similarly sensitive to the statistical style results. In particular, we were interested in whether the subjects and the SOM model are sensitive to certain statistical regularities found in the actual music.

We considered both the distribution of tones and certain sequential aspects of music, in particular, the frequency of two- and three-tone transitions. This approach proved fruitful because listeners seemed to be sensitive to different types of statistical information depending on the level of expertise. It was found that the experts used more detailed information, specifically tone transitions, whereas nonexperts relied more on general principles of music, such as tonality as it is reflected in tone distributions. This extends previous findings by showing not only that behavioral and statistical measures correlate, but that different statistical properties correspond to different levels of expertise. This suggests that it would be desirable to compare behavioral data with more detailed and theoretically motivated statistical analyses of music.

The statistical style analysis of the spiritual folk hymns was also used to investigate the operation of the different SOMs. Recall the Hymn SOM was trained on the hymn fragments used in the behavioral experiment with the correct continuation tones. The Finnish SOM was trained on a set of Finnish folk songs and Lutheran hymns to which the spiritual folk hymns are related. When comparing the response of the two SOMs with the data from the statistical analysis, it was found that the Hymn SOM had a response pattern that more heavily weighted the tone transitions in the hymn corpus than the Finnish SOM. Thus, the same relationships held between the statistical data and both the behavioral and SOM measures, with higher order statistical properties correlating more strongly with the experts and the Hymn SOM. More generally, the correlations between statistical and SOM results implies that the SOM model is sensitive to distributional information and is capable of extracting regular features in the music with which it is presented. This encourages further explorations of the statistical properties to which neural network models are sensitive.

The present study also established a strong relationship between the behavioral responses and the SOM model. This relationship could be seen in the correlations between the responses of the two networks and the two subject groups. The model's responses to musical stimuli were quite similar to the responses of the human subjects. In addition, the two models corresponded to the different levels of listeners' expertise. The experts' responses correlated more strongly with the responses of the Hymn SOM than did the nonexperts' responses. Consequently, using different corpora of melodies for training the model, it was possible to simulate different levels of expertise. This strategy may prove useful in further studies of musical expertise and help to clarify the characteristics of SOM models.

Additional agreement between behavioral and SOM model results was found by analyzing them in terms of the principles of the implication-realization model (Narmour, 1990). The responses of both experts and nonexperts and the Hymn SOM all correlated quite strongly with Proximity, Registral Return, and Intervallic Difference. Low correlations were found with Registral Direction and Closure. Thus, the three sets of data all weighted the principles of the model similarly. Narmour (1990) proposed that the principles are universal and innate, consistent with their similar influences on experts and nonexperts. This similarity across experts and nonexperts was also found previously (Krumhansl, 1995b). However, that the Hymn SOM was similar to both groups of listeners in the present study raises the issue of whether the principles might be abstracted from listening to music. As shown by Thompson and Stainton (1998), music at least in one style tends to conform to the principles. If this is generally true, then it is possible that listeners have abstracted the principles because they describe common patterns in the music they have heard. With the SOM approach, the music presented can be precisely controlled, suggesting that this kind of modeling offers a special opportunity for studying the origin of the proposed bottom-up principles.

The weightings of the principles of the implication-realization model differ somewhat from previous studies (Cuddy & Lunney, 1995; Krumhansl, 1991, 1995a, 1995b, 1997; Thompson et al., 1997; Thompson & Stainton, 1998). Exact comparisons between studies cannot be made, however, for two reasons. The first is that the different studies consider different musical contexts, and these will have different distributions of implicative intervals. For example, one study might have an equal number of large and small implicative intervals, whereas another study (such as the present) has a large majority of small implicative intervals, which are generally more frequent in music. This means that the model's predictions will depend on the particular sequences under analysis. Krumhansl (1997) called this dependency the *contextual distribution* of the principles. The second problem is that different variables have been entered into the analysis together with the implication-realization model's principles. For example, a variable coding tonality for one musical style might be different from another musical style. These differences may affect the results obtained for the implication-realization model's principle in the statistical analysis.

It was possible, however, to test different formulations of the implication-realization model against the present results. The study by Krumhansl (1995a) suggested that consonance was another bottom-up factor that influences melodic expectancies, and this was substantiated by the present results. This finding is interesting in light of the parallel study using Sami yoiks (Krumhansl et al., 1999), in which consonance was the dominant factor. The simplifications of the implication-realization model proposed by Schellenberg (1996, 1997) were not supported. No statistical support was found for the revised principle of Registral Direction, nor for the proposed Pitch Reversal principle. Previous analyses had also shown that, although the revised models contained fewer principles, they performed better than other formulations only on the case of English folk songs. This suggests that it may be premature to reduce the number of factors considered until additional musical styles have been studied.

The SOM model developed its response structure on the basis of recorded musical performances without any external supervision. This suggest that the SOM can be used to emulate two processes of music perception. First, it provides a model of how perceptual schemata of music are developed by passive exposure to musical material. Second, it provides a model of how these schemata are used during music listening for interpreting musical events. Thus, SOM models may provide an important tool for abstracting underlying patterns for further statistical style studies and for generating predictions for perceptual principles. In any SOM simulation, the results obtained depend strongly on both how the data are presented to the map and how the response is interpreted. In other words, the encoding and preprocessing of the input data as well as the postprocessing of the output are important steps of SOM modeling. In the present model, a crucial aspect of the preprocessing is probably how temporal information is represented. In light of the results, it seems that the diffusion memory provides a plausible model of how the temporal dimension is represented in music perception. On the postprocessing side, the fitness measure, defined as the opposite of the quantization error, was found to correlate strongly with subjects' ratings and thus provides an adequate measure of goodness-of-fit.

The model used in the present study is related to those presented by Bharucha and Todd (1989), Gjerdingen (1989, 1990), and Leman (1995). Bharucha and Todd (1989) use a back-propagation network for modeling the learning of musical sequences and the production of musical expectations. There are two main differences between their model and the present one. First, the input to their model consists of chord symbols instead of acoustic information. Second, they use a supervised learning algorithm, whereas the present model is based on self-organization. Whereas the models of Gjerdingen (1989, 1990) are based on a self-organizing architecture (Adaptive Resonance Theory, see Carpenter & Grossberg, 1987, 1990), he also uses a representation derived from musical scores rather than acoustic information. His study is interesting in the present context because the model abstracted common melodic patterns. Perhaps most closely related to the present model is that of Leman (1995; see also Leman & Carreras, 1997). He uses an auditory model and a SOM for modeling the perception of tonality. Starting from acoustic information, he demonstrates that the

SOM develops tonal schemata that have a topographical organization in terms of the circle of fifths.

In summary, we believe a combination of methods, such as those used here, is important for establishing the convergence between different approaches. A recent study (Toiviainen, Tervaniemi, Louhivuori, Saher, Huotilainen, & Näätänen, 1998) on musical timbre found convergence between behavioral judgments of musical timbre similarity, brain wave responses (the magnitude of the mismatch-negativity), and a SOM representation of the timbre space. We adopted a similar strategy in this study. The topic, melodic expectation, was chosen because previous research allowed us to compare the present results with those for other styles and more schematic materials. Theoretical proposals about melodic expectation provided testable predictions about underlying psychological principles, which were evaluated and compared with other empirical tests. The statistical style analysis proved to be an important additional tool for understanding the musical corpus. It provided useful information for selecting the stimulus materials and for numerically representing both general stylistic features and more specific characteristics of the style. Finally, the success of the SOM model supports the adequacy of this type of neural network as a model of learning. Moreover, it demonstrated sensitivity to similar features uncovered in the behavioral experiment and the statistical style analysis, thus establishing strong convergence between all three approaches.[7]

## References


Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5*, 1–30.

Bharucha, J. J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In P. M. Todd & D. G. Loy (Eds.), *Music and connectionism* (pp. 84–99). Cambridge, MA: MIT Press.

Bharucha, J. J. (1999). Neural nets, temporal composites and tonality. In D. Deutsch (Ed.), *The psychology of music* (2nd ed.) New York: Academic Press.

Bharucha, J. J., & Olney, K. L. (1989). Tonal cognition, artificial intelligence and neural nets. *Contemporary Music Review*, *4*, 341–356.

Bharucha, J. J., & Todd, P. M. (1989). Modeling the perception of tonal structure with neural nets. *Computer Music Journal*, *13*, 44–53.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.



7. This research was supported by a Fulbright fellowship, "Music Cognition: Cultural and Neural Substrates," and a faculty grant from Cornell University to Carol L. Krumhansl. It was also supported by the Finnish Academy with the grant, "Tacit Knowledge in Complex Mind Environment Systems" and a grant from the University of Jyväskylä. Topi Järvinen was supported by a postdoctoral research grant from the Finnish Academy, and Tuomas Eerola was supported by a predoctoral research grant from the University of Jyväskylä. We are grateful to the Youth Choir of the Beseechers for participating in the behavioral experiment, and especially to the coordinator of the choir, Markus Halminen, for making the arrangements for the experimental session. Riitta Rautio arranged for the experimental session with the musicology students.

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology, 1*, 12–29.
Carlsen, J. C., Divenyi, P. L., & Taylor, J. A. (1970). A preliminary study of perceptual expectancy in melodic configurations. *Council for Research in Music Education Bulletin, 22*, 4–12.
Carpenter, G. A., & Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics, 26*, 4919–4930.
Carpenter, G. A., & Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks, 3*, 129–152.
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.
Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics, 57*, 451–462.
Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3*, 255–267.
Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.
Gjerdingen, R. O. (1989). Using connectionist models to explore complex musical patterns. *Computer Music Journal, 13*, 67–75.
Gjerdingen, R. O. (1990). Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception, 7,* 339–370.
Griffith, N. (1994). Development of tonal centres and abstract pitch as categorizations of pitch use. *Connection Science, 6*, 155–175.
Halullisten Sielujen Hengelliset Laulut (1963). *Siionin virret ja halullisten sielujen hengelliset laulut* [Hymns of zion and spiritual songs for devoted souls]. Pieksämäki, Finland: Eurajoen Kristillinen Kansanopisto.
Halullisten Sielujen Hengelliset Laulut (1988). *Siionin virret ja halullisten sielujen hengelliset laulut* [Hymns of zion and spiritual songs for devoted souls]. Rauma, Finland: 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., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology, 12,* 154–171.
Järvinen, T. (1995). Tonal hierarchies in jazz improvisation. *Music Perception, 12,* 415–437.
Järvinen, T. (1997). *Tonal dynamics and metrical structures in jazz improvisation.* Published doctoral dissertation. In *Jyväskylä Studies in the Arts.* Jyväskylä, Finland: University of Jyväskylä.
Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum Associates.
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.
Kodály, Z. (1959). A prerequisite condition of comparative song research. In G. Dienes (Ed.), *Studia memoriae Belae Bartók sacra* (3rd ed., pp. 11–12). Budapest: Hungarian Academy of Sciences.
Kohonen, T. (1997). *Self-organizing maps* (2nd ed.). Berlin: Springer-Verlag.
Krohn, I. (1899). *Über die Art und Entstehung der geistlichen Volksmelodien in Finnland.* Unpublished doctoral dissertation, University of Helsinki, Finland.
Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* Oxford: Oxford University Press.
Krumhansl, C. L. (1991). Melodic structure: Theoretical and empirical descriptions. In J. Sundberg (Ed.), *Music, language, speech, and brain.* London: MacMillan.
Krumhansl, C. L. (1995a). Effects of musical context on similarity and expectancy. *Systematische Musikwissenschaft (Systematic Musicology), 3,* 211–250.

Krumhansl, C. L. (1995b). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, *17*, 53–90.

Krumhansl, C. L. (1997). Effects of perceptual organization and musical form on melodic expectancies. In M. Leman (Ed.), *Music, Gestalt, and computing: Studies in cognitive and systematic musicology*. Berlin: Springer-Verlag.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89,* 334–368.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (1999). *Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks*. Manuscript in preparation.

Leman, M. (1994). Schema-based tone center recognition of musical signals. *Journal of New Music Research, 23,* 169–204.

Leman, M. (1995). *Music and schema theory: Cognitive foundations of systematic musicology.* Berlin: Springer-Verlag.

Leman, M., & Carreras, F. (1997). Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In M. Leman (Ed.), *Music, Gestalt, and computing: Studies in cognitive and systematic musicology* (pp. 144–168). Berlin: Springer-Verlag.

Louhivuori, J. (1988). *Veisuun vaihtoehdot: Musiikillinen distribuutio ja kognitiiviset toiminnot* [Variation in spiritual folk hymn singing: Musical distribution and cognitive processes]. Published doctoral dissertation. *Acta Musicologica Fennica 16*. Helsinki, Finland: Suomen musiikkitieteellinen seura.

Merivirta, M. (1930). *Halullisten Sielujen Hengellisten Laulujen ja Siionin Virsien sävelmät.* Jyväskylä, Finland: Jyväskylä Matti Merivirta.

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

Meyer, L. B. (1967). *Music, the arts, and ideas*. Chicago: University of Chicago Press.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model.* Chicago: University of Chicago Press.

Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research, 57*, 103–118.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

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. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, *7*, 109–150.

Schmuckler, M. A. (1990). The performance of global expectations. *Psychomusicology*, *9*, 122–147.

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

Thompson, W. F., Cuddy, L. L., & Plaus, C. (1997). Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Perception & Psychophysics*, *59*, 1069–1076.

Thompson, W. F., & Stainton, M. (1998). Expectancy in Bohemian folk song melodies: Evaluation of implicative principles for implicative and closural intervals. *Music Perception, 15,* 231–252.

Toiviainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, M., & Näätänen, R. (1998). Timbre similarity: convergence of neural, behavioral, and computational approaches. *Music Perception, 16,* 223–241.

Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology, 77,* 206–219.

Unyk, A. M., & Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, *7*, 3–23.

# Appendix A
# Three-Tone Transitions

The tables in this Appendix show all of the three-tone transitions for the corpus of 18 spiritual folk hymns used in the statistical style analysis. Each line in each column has the frequency followed by the tones in the three-tone transition. The transitions that were used in the analysis of the behavioral data are marked with an asterisk (*).

TABLE A1
**Three-Tone Transitions for the Major Spiritual Folk Hymns**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | A | G | 1 | d' | d' | c' | 2 | f | g | a |
| 1 | A | B | c | 1 | e | c | d | 2 | f | g | c' |
| 1 | B | A | A | 1 | e | e | c | 2 | f | g | f |
| 1 | B | A | B | 1 | e | g | a | 2 | g | a | d |
| 1 | B | d | c | 1 | e | g | d | 2 | g | e | f |
| 1 | G | G | c | 1 | e | g | e | 2 | g | f | f |
| 1 | G | c | e | 1 | f | e | a * | 3 | A | G | c |
| 1 | G | g | a | 1 | g | a | a | 3 | G | c | B |
| 1 | a | a | d | 1 | g | c | c | 3 | a | f | f |
| 1 | a | d | f♯ * | 1 | g | c' | b | 3 | b | a | a |
| 1 | a | g | e | 1 | g | d | c | 3 | b | a | g |
| 1 | b | a | b | 1 | g | d | g | 3 | c | B | A |
| 1 | b | c' | b | 1 | g | e | c | 3 | c | c | B |
| 1 | b | c' | d' | 1 | g | g | e | 3 | c | c | d |
| 1 | c | G | g | 1 | f♯ | a | a | 3 | c | d | d |
| 1 | c | c' | b | 2 | B | A | G | 3 | d | d | g |
| 1 | c | d | B | 2 | G | B | d | 3 | d | g | g |
| 1 | c | d | c | 2 | G | c | c | 3 | e | a | g |
| 1 | c | d | g | 2 | G | c | d | 3 | e | d | B |
| 1 | c | f | f | 2 | a | c' | b | 3 | f | e | e * |
| 1 | c | g | f | 2 | a | d | e * | 3 | f | g | g |
| 1 | c' | c' | e* | 2 | a | d | f * | 3 | g | a | b |
| 1 | c' | d' | d' | 2 | a | f | a | 3 | g | a | f |
| 1 | c' | e | g | 2 | a | g | c' | 4 | a | a | g |
| 1 | d | B | A | 2 | b | a | c' | 4 | a | b | c' |
| 1 | d | G | G | 2 | b | a | d | 4 | a | g | g |
| 1 | d | G | c | 2 | b | a | f | 4 | c | c | e |
| 1 | d | c | G | 2 | b | c' | a | 4 | c | e | d * |
| 1 | d | c | c' | 2 | c | c | c | 4 | e | f | d |
| 1 | d | c | d | 2 | c | d | G | 4 | f | a | g |
| 1 | d | c | f | 2 | c | e | a * | 4 | f | d | g |
| 1 | d | c | g | 2 | c | e | e * | 4 | f | e | g * |
| 1 | d | d | G | 2 | c' | a | g | 4 | f | g | e |
| 1 | d | d | d | 2 | d | G | B | 4 | g | e | d |
| 1 | d | e | c | 2 | d | c | B | 5 | d | B | d |
| 1 | d | e | e | 2 | d | d | B | 5 | d | g | f |
| 1 | d | g | a | 2 | d | d | e | 5 | e | e | d |
| 1 | d | g | c | 2 | d | f | a | 5 | e | g | c' |
| 1 | d | g | d | 2 | e | d | G | 6 | B | d | d |
| 1 | d | f♯ | a | 2 | e | d | g | 6 | c | B | c |
| 1 | d' | c' | b | 2 | e | f | f | 6 | e | d | d |

Continued

| | | | |
|---|---|---|---|
| 6 | f | e | c * |
| 6 | g | a | g |
| 7 | c' | c' | b * |
| 7 | d | c | c |
| 7 | d | d | c |
| 7 | e | g | g |
| 7 | g | g | f |
| 8 | B | c | d |
| 8 | c | e | g * |
| 8 | d | e | d |
| 8 | e | c | e |
| 8 | e | f | e |
| 8 | f | f | e |
| 8 | g | c' | c' |
| 9 | c | d | e |
| 9 | e | d | e |
| 10 | g | g | a |
| 11 | f | e | f * |
| 12 | a | g | f |
| 12 | d | e | f |
| 13 | c' | b | a |
| 13 | e | f | g |
| 15 | e | d | c |
| 17 | f | e | d * |
| 25 | g | f | e |

### Table A2
### Three-Tone Transitions for the Minor Spiritual Folk Hymns

| | | | |
|---|---|---|---|
| 1 | B | c | G |
| 1 | B | c | e♭ |
| 1 | B | c | g |
| 1 | G | B | B |
| 1 | G | c | d |
| 1 | a | b♭ | a |
| 1 | a | g | g |
| 1 | a | g | g♭ |
| 1 | b♭ | a | g |
| 1 | b♭ | a♭ | g |
| 1 | b♭ | g | c' |
| 1 | c | G | c |
| 1 | c | d | G |
| 1 | c' | b♭ | a♭ |
| 1 | c' | c' | b♭ |
| 1 | d | G | B |
| 1 | d | c | d * |
| 1 | d | c | e♭ * |
| 1 | d | d | d |
| 1 | d | g | c |
| 1 | e♭ | f | e♭ |
| 1 | f | e♭ | f |
| 1 | f | g | a * |
| 1 | f | g | b♭ * |
| 1 | f | g | c * |
| 1 | g | a | b♭ |
| 1 | g | b♭ | g |
| 1 | g | c | c |
| 1 | g | c' | c' |
| 1 | g | d | f * |
| 1 | g | d | g * |
| 1 | g | g | d * |
| 1 | g | g♭ | g |
| 1 | g♭ | g | g |
| 2 | A | G | g |
| 2 | B | G | G |
| 2 | B♭ | A | G |
| 2 | B♭ | c | d |
| 2 | G | G | c |
| 2 | G | g | f |
| 2 | a | g | d |
| 2 | a | g | f |
| 2 | c | B♭ | A |
| 2 | c | B♭ | c |
| 2 | c | c | e♭ |
| 2 | c | d | c |
| 2 | c | d | f |
| 2 | d | B | B |
| 2 | d | B | G |
| 2 | d | G | c |
| 2 | d | d | B |
| 2 | d | d | g |
| 2 | d | e♭ | c |
| 2 | d | f | f |
| 2 | d | f | g |
| 2 | e♭ | d | G |
| 2 | e♭ | f | a♭ |
| 2 | f | a | g |
| 2 | f | d | e♭ |
| 2 | f | e♭ | c |
| 2 | f | e♭ | e♭ |
| 2 | f | g | d * |
| 2 | f | g | e♭ * |
| 2 | g | d | e♭ * |
| 2 | g | e♭ | c |
| 2 | g | f | a |
| 2 | g | f | d |
| 2 | g | f | f |
| 3 | B | B | c |
| 3 | B | c | c |
| 3 | G | c | B |
| 3 | a♭ | g | c |
| 3 | c | a♭ | g |
| 3 | d | B | c |
| 3 | d | c | a♭ * |
| 3 | d | d | e♭ |
| 3 | d | g | g |
| 3 | e♭ | d | g |
| 3 | g | a | g |
| 3 | g | f | g |
| 3 | g | g | a * |
| 4 | G | c | c |
| 4 | a♭ | b♭ | c' |
| 4 | a♭ | g | e♭ |
| 4 | b♭ | c' | a♭ |
| 4 | c | c | g |
| 4 | c' | a♭ | g |
| 4 | c' | g | a♭ |
| 4 | d | c | B♭ * |
| 4 | e♭ | c | e♭ |
| 4 | e♭ | f | d |
| 4 | f | d | g |
| 4 | f | f | e♭ |
| 4 | g | a♭ | b♭ |
| 4 | g | c | d |
| 4 | g | c' | g |
| 4 | g | e♭ | d |
| 4 | g | f | a♭ |
| 5 | d | e♭ | e♭ |
| 5 | d | f | e♭ |
| 5 | e♭ | d | B |
| 5 | e♭ | d | e♭ |
| 6 | c | g | g |
| 6 | d | g | f |
| 6 | e♭ | d | f |
| 6 | f | a♭ | g |
| 6 | f | g | f * |
| 7 | a♭ | g | f |
| 7 | e♭ | d | d |
| 7 | e♭ | e♭ | d |
| 8 | B | c | d |
| 8 | c | e♭ | d |
| 8 | d | c | c * |
| 8 | e♭ | f | g |
| 8 | g | g | f * |
| 8 | g | g | g * |
| 9 | d | c | B * |
| 12 | c | B | c |
| 12 | c | c | d |
| 14 | d | e♭ | f |
| 17 | d | e♭ | d |
| 18 | g | f | e♭ |
| 23 | f | e♭ | d |
| 25 | c | d | e♭ |
| 32 | e♭ | d | c |

# Appendix B

Figure B1 shows the original quantitative coding of Narmour's implication-realization model (Krumhansl, 1995b). The top of Figure B2 shows Schellenberg's (1996) revised coding of the Registral Direction principle. The bottom of Figure B2 shows Schellenberg's (1997) proposed Pitch Reversal principle.

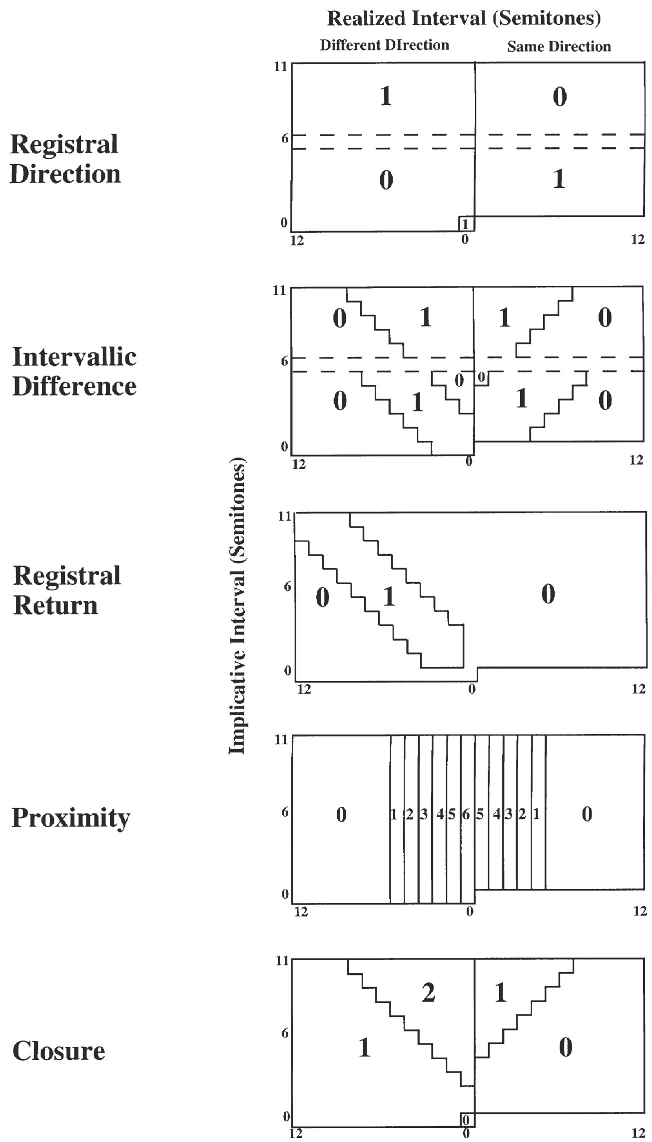


**Fig. B1.** Five principles of the implication-realization model (Narmour, 1990) as originally coded (Krumhansl, 1995b). The principles are Registral Direction (RD), Intervallic Difference (ID), Registral Return (RR), Proximity (PR), and Closure (CL). The vertical axis represents the size of the implicative interval in semitones. The horizontal axis represents the size of the realized interval in semitones and its direction relative to the implicative interval.

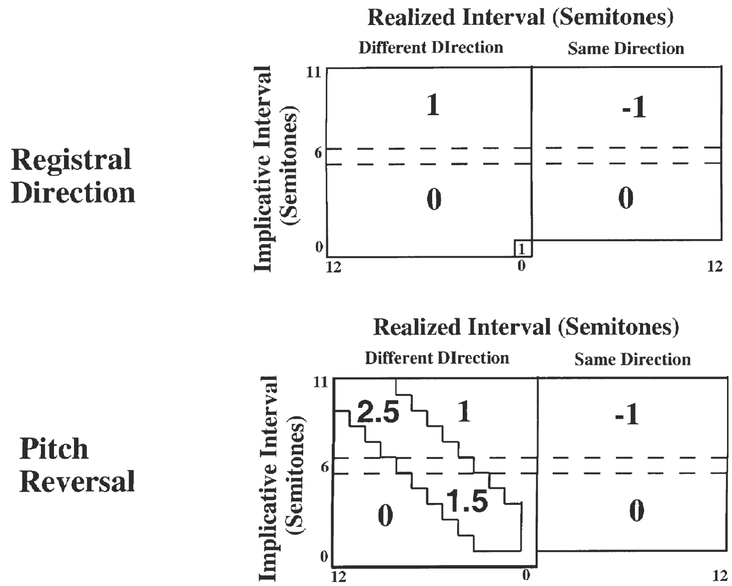


**Fig. B2.** The top grid shows Schellenberg's (1996) proposed revision of Registral Direction. The bottom grid shows Schellenberg's (1997) proposed Pitch Reversal principle, which is a combination of Registral Direction and Registral Return.

# Appendix C

## THE PITCH EXTRACTION ALGORITHM

Let $s(i), i = 0,1,2,\ldots$ denote the series of samples of the sound wave, taken with sampling intervals $\Delta t$. The periodicity analysis is carried out by a set of autocorrelation units, where each unit calculates the exponential average of the product of two signals—the actual input signal and the same signal shifted backward in time by

$$\Delta t_j = 2^{-j/12}\,\Delta t_0,\ j = 0, \ldots, n, \tag{1}$$

The time lags $\Delta t_j$ are thus logarithmically spaced, and correspond to semitone intervals; each time lag $\Delta t_j$ corresponds to frequency $f_j = 1/\Delta t_j$. The activation value of autocorrelation unit $j$ at time $i$ is calculated by

$$a_j(i) = e^{-\Delta t/\tau} a_j(i-1) + (1 - e^{\Delta t/\tau}) s(i) s(i - \Delta t_j/\Delta T), \tag{2}$$

where $\Delta T$ denotes the sampling interval of the sound file and $\tau$ the time constant (the value $\tau = 20$ ms was used).

The states of the autocorrelation units are sampled at $t_k = k\Delta\tilde{T}$, where $\Delta\tilde{T} = 40$ ms. The thus obtained vector $\mathbf{a}(t_k)$ is a half-wave rectified according to

$$b_i(t_k) = \begin{cases} a_i(t_k), \text{ if } a_i(t_k) \geq 0 \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

Vector $\mathbf{b}(t_k)$ represents the instantaneous pitch as peaks at the fundamental frequency and its subharmonic series. To remove the subharmonic components, $\mathbf{b}(t_k)$ is multiplied by matrix $\mathbf{A} = (a_{ij})$, defined by

$$a_{ij} = \begin{cases} 1, \text{ if } i = j \\ -1, \text{ if } j - i \in \{12,19,24,28,31,...\}; \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

the result is denoted by $\mathbf{c}(t_k)$. The pitch of the input signal at $t_k$ is now represented by vector $\mathbf{d}(t_k)$, obtained by

$$d_j(t_k) = \begin{cases} c_j(t_k), \text{ if } j = \arg\max_{\eta} c_\eta(t). \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

In other words, only the maximum component of each $\mathbf{c}(t_k)$ is retained while the other components are set to zero.

In the experiments, 22 autocorrelation units were used; their respective frequencies ranged from $G_3$ to $E_5$, corresponding to the lowest and the highest pitch in all 18 hymns.

## DIFFUSION MEMORY

Using the pitch vectors $\mathbf{b}(t) = (b_0(t),\ldots,b_n(t))$ as input, a memory trace is built by mapping the temporal dimension to a spatial dimension by means of a diffusion mechanism. The memory trace is represented by a matrix $\mathrm{C} = (\mathrm{c}_{kj})$, $k = 1,\ldots,m$, $j = 1,\ldots,n$, whose time evolution is determined by equations

$$\begin{cases} \dot{c}_{1j} = b_j - \alpha \mathrm{c}_{1j} \\ \dot{c}_{kj} = \alpha(c_{(k-1)j} - c_{kj}),\ k = 2,\ldots,m \end{cases} \tag{4}$$

In the discrete domain, Equations 4 are approximated by

$$\begin{cases} c_{1j}(t) = c_{1j}(t-1) + \Delta\tilde{T}(b_j(t-1) - \alpha c_{1j}(t-1)) \\ \mathrm{c}_{kj}(t) = \mathrm{c}_{kj}(t-1) + \Delta\tilde{T}\alpha(c_{(k-1)j}(t-1) - \mathrm{c}_{kj}(t-1)),\ k = 2,\ldots,m \end{cases} \tag{5}$$

Here $\dot{c}$ stands for the time derivative of $c$, $\alpha$ for the rate of diffusion, and $\Delta\tilde{T}$ for the sampling interval between successive $\mathbf{b}(t)$. Each pitch is thus represented with a chain of components $(c_{1j},\ldots c_{mj})$, where the activation flows from $c_{1j}$ through the chain toward $c_{mj}$. In the simulations, the length of the diffusion chain was set to $m = 6$.

Starting from Equations 4, it is straightforward to show that, after $c_{1j}$ receives an input, the time it takes for $c_{jk}$ to attain its maximal activation level is $m/\alpha$. In the text, this time is referred to as the memory length. It should be noted, however, that the memory trace is actually longer than $m/\alpha$; after the last unit has attained its maximal activation level, it starts to decay. Under the absence of any further input, the decay is approximately exponential.