Music Perception
Fall 2002, Vol. 20, No. 1, 51–85

Harmonic Factors in the Perception of Tonal Melodies

DIRK-JAN POVEL & ERIK JANSEN
Nijmegen Institute for Information and Cognition

By common assumption, the first step in processing a tonal melody consists in setting up the appropriate metrical and harmonic frames required for the mental representation of the sequence of tones. Focusing on the generation of a harmonic frame, this study aims (a) to discover the factors that facilitate or interfere with the development of a harmonic interpretation, and (b) to test the hypothesis that goodness ratings of tone sequences largely depend on whether the listener succeeds in creating a suitable harmonic interpretation. In two experiments, listeners rated the melodic goodness of selected sequences of 10 and 13 tones and indicated which individual tones seemed not to fit. Results indicate that goodness ratings (a) are higher the more common the induced harmonic progression, (b) are strongly affected by the occurrence and position of nonchord tones: sequences without nonchord tones were rated highest, sequences with anchoring nonchord tones intermediately, and nonanchoring nonchord tones lowest. The explanation offered is compared with predictions derived from other theories, which leads to the conclusion that when a tone sequence is perceived as a melody, it is represented in terms of its underlying harmony, in which exact pitch-height characteristics play a minor role.

Received September 20, 2001, accepted December 31, 2001.

THIS study is about the perception of tonal melodies and particularly the process by which the notes of simple tonal melodies are transformed into a musical percept. Numerous theoretical and empirical studies, some of which are discussed in detail here, have indicated the crucial importance of the development of an appropriate frame of reference, or schema, that identifies the context in which the tones in the input are interpreted and that guides further processing, possibly leading to a musical representation (e.g., Bharucha, 1987, 1991; Cuddy, Cohen, & Mewhort, 1981; Gjerdingen, 1990; Holleran, Jones, & Butler, 1995; Krumhansl, 1990; Krumhansl & Kessler, 1982; Lerdahl, 1988; Lerdahl & Jackendoff, 1983; Longuet-Higgins

Address correspondence to Dirk-Jan Povel, NICI, P.O. Box 9104, 6500 HE Nijmegen. The Netherlands. (e-mail: povel@NICI.kun.nl)
ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

& Steedman, 1971; Povel, 1981; Povel & Essens, 1985; Sloboda & Parker, 1985; Thompson, 1993; Tillmann, Bharucha, & Bigand, 2000; Van Dyke Bingham, 1910). For tonal music, this frame of reference consists of a metrical frame that enables the interpretation of the rhythmical aspects and a harmonic frame that allows the interpretation of the melodic/harmonic aspects.

The present study focuses on the induction of the harmonic frame and is based on the assumption that a sequence of tones will be perceived as a tonal melody only if the listener succeeds in discovering an acceptable underlying harmony. By comparing the characteristics of sequences that are perceived as good or bad melodies, we want to discover the factors that determine whether the listener succeeds in creating an appropriate harmonic analysis and to develop hypotheses about how these sequences are processed. Before dealing with the investigation proper, we discuss a few earlier studies that have greatly influenced the approach taken in the present study.

Van Dyke Bingham (1910) performed an experiment in which all melodic intervals within one octave were presented, and subjects answered the question "Can you make this second tone a final tone? Does this melody end?" (p. 23). The results indicate that the descending perfect fifth, the descending major third, and the ascending perfect fourth show the strongest tendency to be heard as final. From a detailed analysis of all his results he concluded that: "Two melodically 'related' tones tend to establish a tonality, and the melody [the melodic interval] is judged to end only when the final tone is one of the members of the tonic triad—preferably the tonic itself" (p. 34). This is an interesting observation that indicates that even when perceiving single melodic intervals listeners tend to interpret the tones in a tonal frame by establishing a key.

Sloboda and Parker (1985) reported an exploratory study in which eight subjects, four with musical training and four without, provided six successive sung recalls of a fragment of the Russian folksong "Sailor" (comprising 30 notes). After transcription, the reproductions were analyzed in several ways including a melodic contour analysis, a metrical analysis, a rhythmical analysis, a phrase structure analysis, and a harmonic analysis. The main findings were that (a) recall of the melody was never perfect, even for the musically trained subjects, (b) the metrical structure is always preserved, suggesting that meter is a primary structural frame for melodic recall, but the actual rhythms were in about half of the cases substituted by metrical equivalents, (c) the harmonic structure is coded but the exact melodic structure is often lost, (d) musicians and nonmusicians appear to process the music in much the same way, except for the harmonic relationships, which are better coded by the musicians, (e) subjects' performance did not improve during the six trials on any of the measures.

The authors summarize the result of their study as follows: "memorizing simple, well-formed tonal melodies involves building a mental model of the underlying structure in which not all of the surface detail is necessarily retained." (p. 160). Thus, even when subjects are explicitly asked to reproduce a melody, they appear to be unable to make a literal reproduction, but rely on the generated underlying structure consisting of a metrical frame and a harmonic frame to generate a paraphrase. The finding that the subjects' performance did not improve during the six trials is quite surprising and may indicate that the exact coding of the melody is not part of the everyday listening process.

Research by Holleran, Jones, and Butler (1995), Platt and Racine (1994), and Thompson and Cuddy (1989) confirm the results of the study by Sloboda and Parker (1985) by providing experimental evidence that listeners make a harmonic analysis if the input consists of a single voice melody. Trehub, Thorpe, and Trainor (1990) reported that infants 7–10 months old can detect small changes to tone patterns based on a V-I chord progression better than changes to patterns not based on such a chord progression, suggesting that even these very young children are sensitive to the underlying harmony (see, however, Trainor & Trehub, 1992).

Cuddy, Cohen, and Mewhort (1981) studied the perception of tone sequences having "varying degrees of musical structure." They constructed a set of sequences by altering one or more tones of the "prototypical" sequence C_5 E_5 G_5 F_5 D_5 B_4 C_5 (the numeral indicates the octave; C_4 = middle C) thereby gradually distorting the "harmonic structure," contour complexity, and excursion size (interval between first and last tone). Based on the results of Experiment 1, in which subjects judged the "tonality or tone structure" of 32 seven-tone sequences, five levels of harmonic structure were constructed by combining three rules: (1) diatonicism (a series may or may not consist of only diatonic tones); (2) leading-note-to-tonic ending; (3) the extent to which a sequence follows a I–V–I harmonic progression. These levels of harmonic structure were factorially combined with two levels of contour complexity and two levels of excursion, yielding 20 stimuli that were recognized under transposition (Experiment 2) and rated on tonal structure (Experiment 3). Findings indicate that the ratings were mostly influenced by the factor harmonic structure and less by contour and excursion.

Because the latter study has greatly influenced the research presented here, we discuss a few of its aspects in more detail. First, the concept of harmonic structure, as expressed in the five levels of harmonic structure, is not theoretically but empirically determined. As a result, it is unclear how the three rules have precisely determined the variable harmonic structure. Second, although the rules describe listeners' responses to the 20 sequences of Experiments 2 and 3 reasonably well, it is unclear to what extent the rules can be generalized to other tone sequences. This is clarified with the melodic examples in Figure 1.

Although Sequences 1a and 1b in Figure 1 both violate the rule of diatonicism, Sequence 1b (containing two chromatic tones) will receive a higher goodness rating than Sequence 1a, most likely because the F♯ and D♯ respectively resolve to the succeeding G and E. Sequences 2b and 2c both violate the leading-tone-to-tonic-ending rule, but they will probably be rated equally high. Finally, Sequences 3b and 3c do not have an underlying I–V–I progression, but they will still be judged good melodies or melodic fragments. These examples do not undermine the general finding that harmonic

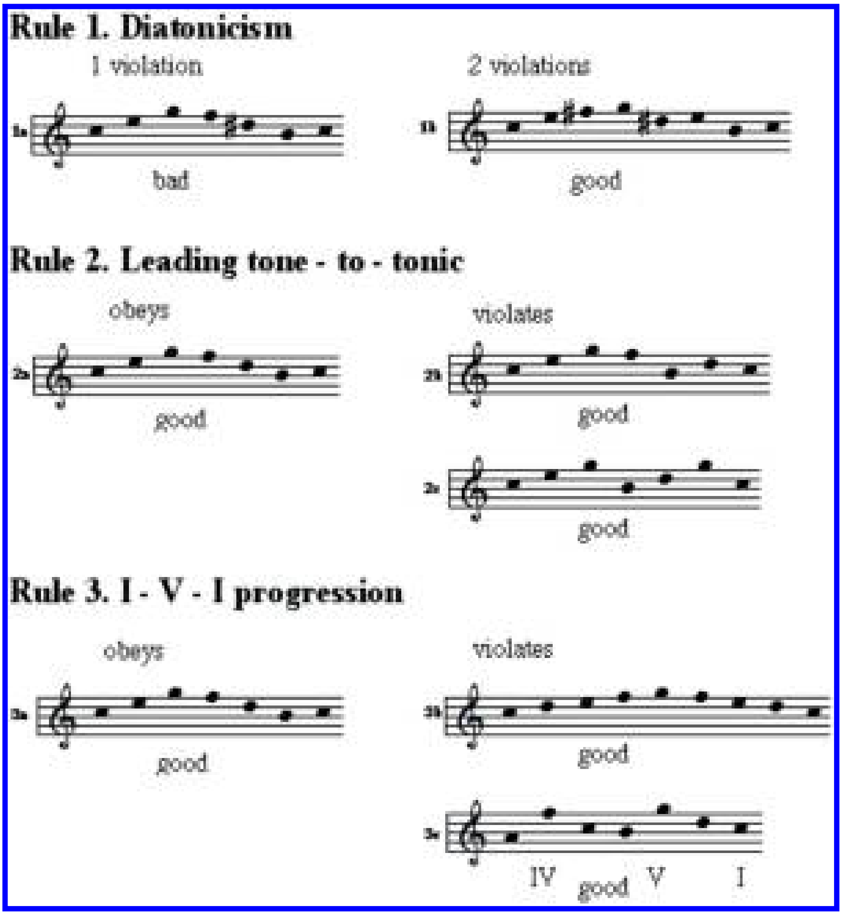

Fig. 1. Examples of sequences that do not obey the harmonic rules of Cuddy, Cohen, and Mewhort (1981), but still form good melodies. 1a. This sequence violates the diatonicism rule and indeed sounds bad. 1b. This sequence contains 2 violations of the diatonicism rule, yet it sounds good. Sequences 2b and 2c both violate the leading tone-to-tonic rule, yet they are good melodies. Sequences 3b and 3c both do not follow a I–V–I chord progression, still they are good melodies.

structure is a major factor in the perception of tone sequences, but they indicate that their definition of harmonic structure is still incomplete.

Next, we want to advance some speculative explanations for why some sequences from the Cuddy et al. (1981) study were judged as bad melodies. Some of the stimuli may have been rated low merely because they are experienced as unfinished. For example, the sequence E_5 B_5 G_5 F_5 D_5 A_5 C_5 (Stimulus 23, Expt. 1, rating 2.9 on a 6-point scale[1]) can be transformed into a good sequence by adding a few tones: E_5 B_5 G_5 F_5 D_5 A_5 C_5 G_4 B_4 C_5. In a similar fashion, the sequence C_5 E_5 B_5 G_5 F_5 D_5 A_5 (Stimulus 18, Expt. 1, rating 3.2) can be transformed into a good one by extending it as follows: C_5 E_5 B_5 G_5 F_5 D_5 A_5 G_5 F_5 B_4 C_5. The incompleteness of these two sequences thus seems to have been caused by specific implications, created by the succession of tones, that are not met. Note that both sequences contain the fragment B_5 G_5 F_5 D_5, which is probably recognized as a G7 chord, implying a solution to elements of the tonic, preferably the C, which does appear in the completed sequences. Still another example is Stimulus 11 (rating 4.4), C_5 E_5 G_5 F_5 D_5 B_4 A_4. This sequence can very simply be altered to form a good melody by adding either B_4 C_5 or G_4 C_5. The reason that this sequence received a higher rating than the previous two may be that only two tones are needed to complete the sequence, whereas in the former ones more tones are needed for completion, thereby putting more strain on memory and imagination. The same reasoning may be applied to Stimulus 22 (rating 3.0): D_5 F_5 $G\sharp_5$ $F\sharp_5$ $D\sharp_5$ B_4 C_5. The fragment $G\sharp_5$ $F\sharp_5$ $D\sharp_5$ induces the G♯7 chord, creating an expectation for elements of the tonic C♯ but the tones B and C do not fit these at all.

The picture that arises from these descriptions is that the listener in his/her attempt to create a musical representation encounters a fragment that induces a musical interpretation (e.g. a V7 chord), and next determines whether the following tones somehow fit with the expectation(s) created by that interpretation. If this happens to be the case, the sequence will be judged a good melody, otherwise perception will fail and the sequence will be judged a bad melody.

Finally we consider Stimulus 13 (rating 4.3) C_5 E_5 G_5 $F\sharp_5$ ($G\flat_5$) D_5 B_4 C_5. This sequence can be repaired in two ways by invoking anchoring: C_5 E_5 G_5 $G\flat_5$ F_5 D_5 B_4 C_5 or C_5 E_5 G_5 $G\flat_5$ F_5 D_5 B_4 G_4 C_5. The second sequence seems to sound better probably not because of harmonic but of rhythmical reasons. The first sequence, since it consists of eight tones, cannot be given a metrical interpretation such that it ends on a downbeat. The second sequence however, consisting of nine tones, can be conceived in a metrical frame, by placing downbeats on the first, fifth, and last tone. Thus the last sequence ends on a downbeat and is metrically well-formed.

1. The rating of the highly trained subjects is given here.

From these examples, it may be concluded that the goodness of a sequence is determined both by harmonic and metrical factors: the harmonic interpretation must be such that implications are resolved, and the metrical interpretation should allow the organization of the rhythmical content. These ideas will be elaborated on further later.

Povel and Jansen (2001) and Jansen and Povel (2000) performed a series of studies in which listeners judged the goodness of different sets of tone sequences: (a) tone sequences containing both diatonic and chromatic tones, (b) tone sequences containing only diatonic tones, and (c) tone sequences containing only arpeggiated chords. These studies are reviewed in Povel and Jansen (2000). Povel and Jansen (2001) studied the perception of a set of tone sequences consisting of a subset of all orderings of the collection C_4 E_4 $F\sharp_4$ G_4 $B\flat_4$. Each sequence was preceded by the chords C7–F to induce the key of F major. It was hypothesized that a tone series is judged a good melody if either one or both of the perceptual mechanisms *chord recognition* and *anchoring* (Bharucha, 1984, 1996) can be applied to the series. Chord recognition is the mechanism that describes a series of tones as a chord, and anchoring is the mechanism that links a tone to a (chord) tone occurring later in the series. Applying these mechanisms, a sequence of tones is conceived as the chord C7, if the $F\sharp_4$, which does not belong to the chord, can be "anchored" to a subsequent G_4. Anchoring may either be immediate when the G follows the F♯, as in the tone series C_4 E_4 $F\sharp_4$ G_4 $B\flat_4$, or more or less delayed when one or more tones intervene between the F♯ and G, as in the series E_4 $F\sharp_4$ C_4 G_4 $B\flat_4$ or $B\flat_4$ $F\sharp_4$ E_4 C_4 G_4. Experiments in which listeners rated melodic goodness and produced the expectations created at different positions in the sequence supported the hypothesis. The same hypothesis was tested for sequences containing only diatonic tones, notably a subset of 60 sequences from the set containing all orderings of the collection D_4 E_4 F_4 G_4 A_4 B_4 (Jansen & Povel, 1999). Although the responses showed considerable interindividual differences, the results globally supported the hypothesis.

Jansen and Povel (2000) studied sequences consisting of arpeggiated chords. Thirty-two six-tone sequences were constructed, each metrically segmented into two groups of three. Each group consisted of tones from one of the triads I, IV, or V. Chord progressions were formed by four different combinations of these triads. Contour complexity of the sequences was also manipulated. Ratings of the tone sequences indicate that the goodness responses are determined by the usualness of the perceived implied harmonic progression (e.g., I–IV being rated higher than V–IV), as well as by the contour complexity of the sequences.

THE PROCESSING OF SIMPLE TONAL MELODIES

If we combine the results of the studies just discussed, the following conceptualization of the processing of tonal melodies emerges. Upon hearing the initial tones of a sequence, the listener attempts to establish the interpretational context of the sequence: the metrical frame and the harmonic frame. As this study focuses on the induction of the harmonic frame, we do not discuss the metrical aspect of the interpretational context here (all tone sequences used in the experiments induce the same meter). The harmonic frame has a global and a local aspect: the global context consists of a key and a mode, whereas the local context consists of a region within the key (I, V, vi, etc). After the listener has established key and mode, (s)he will attempt to divide the sequence into regions, each associated with a harmony. How the processes of key finding and region assignment precisely interact is not well understood. Although the system is usually conceived as completely hierarchical (e.g., Bharucha, 1987; Tillmann, Bharucha, & Bigand, 2000), some authors have argued for a partially hierarchical system (Povel & Van Egmond, 1993; Thompson, 1993; Thompson & Cuddy, 1989). The harmonic function of the region determines the musical function of the tones within that region, thereby determining the stability of the tones and setting up expectations for resolutions of unstable tones. The regions themselves also differ in stability and also create expectations for succeeding regions. Thus the harmonic analysis has two levels: on the highest level functions the key, on the lower level the region.

Whether the listener succeeds in making a harmonic analysis largely depends on the distribution of the tones and especially on the occurrence and localization of nonchord tones (nonchord tones are of course only defined if surrounding tones are interpreted as chord tones). Nonchord tones indeed form a central problem in the development of an algorithmic harmonic analysis (Temperley, 1997). Nonchord tones can be incorporated in a harmonic analysis (a) if they can be linked to a following chord tone (anchoring) or (b) if they are assimilated in a run of steps (a series of minor or major seconds). If the harmonic analysis has been successful (meaning that nonchord tones are somehow accommodated), the listener determines whether the chordal implications of the activated regions are resolved (which will generally be the case if the harmonic progression follows Piston's table [Piston & Devoto, 1989] of usual root progressions). It is assumed that only as long as the metrical and harmonic interpretation is successful is complete processing of the sequence (e.g., the coding of the actual rhythmical figures and pitch patterns) possible.

The ideas proposed here are illustrated with a few melodic examples (Figure 2). The supposed goodness (or badness) of these sequences will be explained in terms of these ideas. The fact that the A in Sequence *a* does not seem to fit can be explained in two ways: (a) the A does not fit in the G7 chord, which is induced by the G F and D (which would be resolved by the subsequent C); (b) upon hearing the A, the tones F D A induce the region ii, resulting in the progression I–ii–I, which is rather unusual. These two hypotheses are put to the test in the next melodic examples. In Sequence *b*,

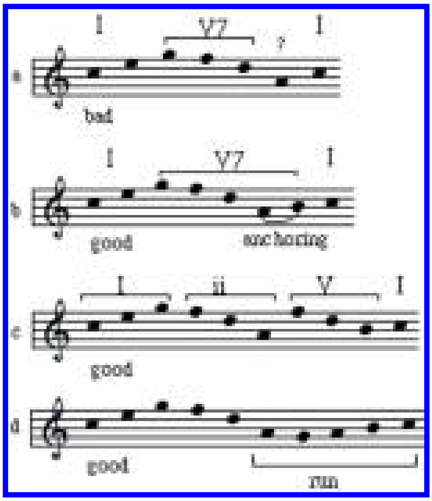

Fig. 2. Harmonic factors that determine melodic goodness (see text for explanation).

the A is incorporated in the G7 chord by means of the added tone B, to which the A can be anchored. In Sequence *c*, the ii chord is followed by a V7 chord, yielding the common progression I–ii–V7–I. Sequence *d*, finally, shows how the A is incorporated in a scale fragment or run of steps.

THE PRESENT RESEARCH

On the basis of the studies just described and the emanating theoretical considerations proposed, we can now list the major reasons why a listener judges a tone sequence as not being a good melody. These are (a) the sequence is metrically indeterminate: the listener is unable to make a suitable metrical interpretation, for instance because the sequence does not end on a downbeat , etc., (b) the sequence is harmonically indeterminate: the listener is unable to make a coherent harmonic interpretation, (c) the sequence induces an implied harmony with an irregular harmonic rhythm, (d) the sequence is not finished, that is, the perceptual processing of the sequence has created one or more expectations (implications) that remain unresolved, (e) the contour of the sequence is relatively complex (Cuddy et al., 1981, Jansen & Povel, 1999), and (f) the first and last tone of the sequence are not the same (Cuddy et al., 1981).

Earlier research on rhythm perception showed that the discovery of an underlying meter is a prerequisite for the formation of an accurate mental representation of a rhythmical pattern (Essens & Povel, 1985; Povel, 1981; Povel & Essens, 1985). Analogously, in the present research we focus on the harmonic interpretation and test the hypothesis that a tone sequence can lead to a musical percept only if the listener succeeds in discovering the underlying harmonic frame. The main question we hope to answer is what the conditions are for a successful harmonic analysis. The exact hypothesis reads: A tone sequence will lead to a musical percept only if: (a) the attempt to make a harmonic analysis succeeds and (b) the induced harmonic progression consists of a succession of regions in which the harmonic expectations are resolved. The predictions of these hypotheses are schematically represented in Figure 3.

COMPARISON WITH ALTERNATIVE EXPLANATIONS

The explanation just proposed, which capitalizes on the creation of a harmonic frame in the process of developing a musical representation, will be compared with three alternative explanations, notably: (a) predictions derived from the tonal hierarchy model of Krumhansl (1990; Krumhansl & Kessler, 1982); (b) predictions derived from the theory of Narmour (1990); and (c) predictions based on specific features in the sequence.

Predictions Derived from the Tonal Hierarchy Model

According to the tonal hierarchy model, the tones in a key differ in their "stability": the tonic being most stable, followed by the other tones of the tonic triad, the diatonic tones, with the nondiatonic tones being the least stable. This variable, called "tonality," has been investigated in studies by Cuddy and Lunney (1995), Krumhansl (1995), and Schellenberg (1996, 1997) and shown to play a role in listeners' ratings of how well a third tone continues a two-tone sequence. On the assumption that a sequence will form a better melody as its tones have a higher average stability, we have

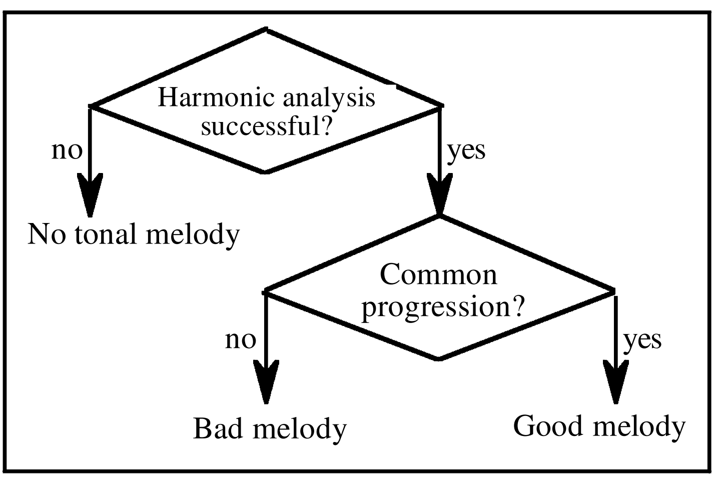

Fig. 3. Schematic representation of the predictions made in Experiment 1.

calculated a variable *MeanStab,* which is the mean of the stabilities of the tones in the sequence.

Predictions Derived from Narmour's Theory

Narmour's implication-realization model (Narmour 1990, 1992), assumes that the expectations listeners form when they hear a melody are based on a limited number of factors that are partly innate and partly learned. Here we focus on the innate factors, which are related to the Gestalt principles of proximity, similarity, and symmetry. Central in the implication-realization model are the concepts of an *implicative interval* (consisting of two tones) followed by a *realized interval* (consisting of the last tone of the implicative interval and the following tone). According to the theory, an implicative interval implies that some tones are more likely to follow than others, which signifies that some successions of implicative and realized intervals (strings of three tones) are more expected than others. One component of the implication-realization model, namely that which deals with the innate expectations created by contour characteristics, was formalized by Schellenberg (1996) in terms of five variables: *registral direction*, *intervallic difference*, *registral return*, *proximity*, and *closure*.

The variables are described here only briefly; for a more detailed description see Krumhansl (1995) or Schellenberg (1996). The first two principles, which form the core of the theory, depend on the size of the implicative interval: small (5 semitones or less) or large (7 semitones or more). The tritone (6 semitones) is considered a threshold interval, being neither small nor large.

1. The *principle of registral direction* states that a small implicative interval implies a continuation of pitch direction, whereas a large interval implies a change of direction.
2. The *principle of intervallic difference* states that a small implicative interval implies a similarly sized realized interval whereas a large implicative interval implies a relatively smaller realized interval.
3. The *principle of registral return* describes an archetypical melodic form, namely, the tendency for the second tone of the realized interval to be proximate in pitch to the first tone of the implicative interval (thus forming a symmetric, or approximately symmetric, melodic figure such as C_4 E_4 C_4, or C_4 E_4 $C\sharp_4$).
4. The *proximity principle* describes a general expectancy for small realized intervals.
5. The *principle of closure* depends on pitch direction and interval size. The degree of closure is higher if there is a change in pitch direction and if the realized interval is smaller than the implica-

tive interval. Closure is also dependent on other, style-specific, factors such as duration of the involved tones, position in the metric frame, and harmony.

These variables don't do justice to the comprehensiveness of the implication-realization model, as they don't sufficiently reckon with long-term effects and context effects due to harmonic and metrical factors.

The principles were quantified and represented using a grid representation by Krumhansl (1995, p. 73), and Schellenberg (1996, pp. 78–79) and tested on three-tone sequences (Cuddy & Lunney, 1995), and on the last three tones from initial fragments of British folk songs, atonal melodies, and Chinese folk songs (Krumhansl, 1995; Schellenberg, 1996, 1997) by asking listeners to rate how well the last tone continued the melodic fragments. In the latter investigations, the first tone of the implicative interval always had a longer duration, a stronger metrical position, and was more stable in the key of the fragment than the second tone of the interval. These methodological procedures imply that possible effects from metrical and harmonic factors will be minimal. Together these studies indicate that the principles are indeed valid predictors of listeners' expectations, together with the aforementioned variable "tonality." It may thus be concluded that the principles that basically concern pitch height configurations play an important role in determining listeners' expectancies when perceiving short musical fragments.

Both Cuddy and Lunney (1995) and Schellenberg (1996) noted that the original principles differed considerably in their predictive power and proposed a modification of the model. In the first instance, Schellenberg (1996) reduced the model to three principles but later showed that the model could be reduced to only two principles without any loss of explanatory power (Schellenberg, 1997). These two principles are as follows:

1. The *proximity principle,* which states that "when listeners hear an implicative interval in a melody, they expect the next tone to be proximate to the second tone of the implicative interval (i.e., they expect a small realized interval)" (Schellenberg, 1997, p. 309). This is the same as saying that a listener when listening to a melody at any time expects a small interval. The principle is quantified by assigning a value of 0 to the interval 0 (prime), 1 to a minor second, and so on.
2. The *pitch reversal principle*. This principle is a linear combination of the (revised) principles of *registral direction* and *registral return*. The two principles are virtually uncorrelated (r = .029). The quantification of the two principles is shown in Figure 4.

As the implication-realization theory basically is a bottom-up model that treats melody primarily as a note-to-note phenomenon, as reflected in the

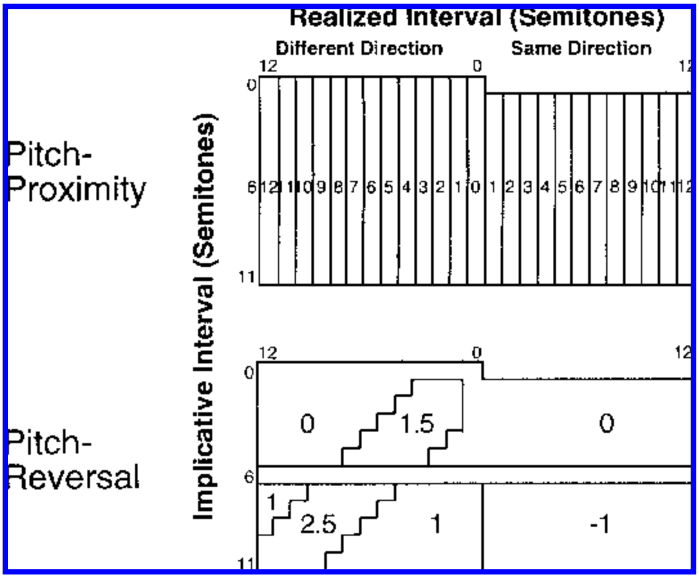

Fig. 4. Quantification of the principles of pitch proximity and pitch reversal proposed by Schellenberg (1997). Figure reprinted with the kind permission of the author and the Regents of the University of California.

principles just discussed, it cannot directly be applied to the perception of entire sequences. Nevertheless, we have computed derivations of the aforementioned principles applicable to an entire sequence by conceiving a tone sequence as a series of consecutive implication-realization intervals (see Thompson & Stainton, 1998). The exact computation of the two modified principles, called *MeanInt* and *PitchRev* is described in the Methods section.

Predictions Based on Specific Sequence Characteristics

Finally, we have measured a few attributes of the sequences that may affect goodness ratings: the *number of contour changes*, the *average consonance of the intervals* in the sequence, and *the variation in interval sizes*.

Experiment 1

In this experiment, we compare the perception of sequences that (according to the notions just considered) possess features that interfere with the generation of a harmonic interpretation, with sequences that lack these features. The “standard” sequences contain tones in positions that hamper a smooth harmonic interpretation, whereas in the “derived” sequences, these tones are replaced or placed in a context that supposedly eliminates or reduces their interference effect. The derived sequences are of three types: (1) sequences consisting of arpeggiated chords, (2) sequences containing nonchord tones that can be linked to chord tones by using the anchoring mechanism, and (3) sequences in which nonchord tones are incorporated in runs. It is predicted that sequences complying with the hypothesized characteristics, and that therefore should enable a straightforward harmonic analysis, will yield a higher goodness rating than sequences that lack these features. To obtain a more detailed insight into which tones actually interfere with the processing of the sequences, participants also rated the individual tones of the stimuli, separately from the global ratings they provided.

METHOD

Subjects

Forty-eight listeners participated in the experiment, consisting of two groups: 22 subjects from the Nijmegen university community (median age = 28.5 years), all of whom were at least moderately trained in music (mean = 10.2 years, *SD* = 7.3 years), and 26 students from two music cognition classes at Northwestern University (median age = 21.5 years) all with a considerable amount of musical training (mean = 15.2 years, SD = 4.7 years).

Construction of the Stimuli

The 20 stimuli used in the experiment, all 10-tone sequences, are displayed in Table 1. We started with five standard sequences, four of which were used in Experiments by Cuddy, Cohen, and Mewhort (1981), selected because of the low ratings they had received in that study. To each of these sequences, 3 tones were added, without changing the harmonic context, so as to make them 10 tones long. These standard sequences contain tones that appear to hamper the creation of a harmonic analysis.

From each of these standard sequences, three new sequences were derived by altering one or more tones. Three types of transformation were used: (1) A transformation yielding a sequence consisting only of chord tones, that is, arpeggiated chords, denoted *harmonic* in Table 1; (2) A transformation yielding a sequence containing one or more nonchord tones that can be linked to a subsequent chord tone by means of the mechanism of anchoring, denoted *anchoring* in Table 1; (3) A transformation yielding a sequence containing one or more nonchord tones that are incorporated in a run (a series of tones with intervals of a semitone or whole tone), denoted *run* in Table 1. The transformations were achieved by applying a minimal number of alterations. Because we wanted to avoid repetition of the same stimulus and because in the *anchoring* and *run* transformations all nonchord tones had to be accommodated by these principles, in a number of cases the derived sequences differ quite a bit from the standard sequences. However, for the question under consideration, this difference is of no consequence.

To control for the temporal grouping or segmentation, all sequences were presented with a distinct timing and dynamic pattern that induced a triple grouping, such that the first and the last tone fell on a downbeat. This timing and accentuation pattern was obtained by recording the performance of a 10-tone sequence presented in a 9/8 meter. The average interonset interval between the tones of this template was 600 ms, whereas the velocity (intensity) of the first tones of the groups (downbeats) was approximately 60 (on a scale from 1to 127), that of the second tone 38, and that of the last tone 33. All stimuli were generated with this timing and accentuation pattern.

TABLE 1
Tone Sequences Used in Experiment 1

No.	Type	Sequence
1	**Standard[a]**	$\mathbf{C_5\ E_5\ G_5\ E_5\ C_5\ G_5\ F_5\ D_5\ A_4\ C_5}$
2	Harmonic	$C_5\ E_5\ G_5\ F_5\ D_5\ A_4\ D_5\ B_4\ G_4\ C_5$
3	Anchoring	$C_5\ E_5\ G_5\ F_5\ D_5\ A_4\ B_4\ G_4\ D_5\ C_5$
4	Run	$C_5\ E_5\ G_5\ F_5\ D_5\ A_4\ B_4\ C_5\ D_5\ C_5$
5	**Standard[b]**	$\mathbf{C_5\ E_5\ G_5\ E_5\ G_5\ C_5\ F_5\ D\sharp_5\ B_4\ C_5}$
6	Harmonic	$C_5\ E_5\ G_5\ E_5\ G_5\ C_5\ F_5\ D_5\ B_4\ C_5$
7	Anchoring	$C_5\ E_5\ G_5\ F_5\ D\sharp_5\ D_5\ B_4\ G_4\ B_4\ C_5$
8	Run	$C_5\ E_5\ G_5\ F_5\ D\sharp_5\ D_5\ C_5\ B_4\ D_5\ C_5$
9	**Standard**	$\mathbf{C_5\ E_5\ G_5\ E_5\ G_5\ C_5\ F\sharp_5\ D\sharp_5\ B_4\ C_5}$
10	Harmonic	$C_5\ E_5\ G_5\ E_5\ G_5\ C_5\ F\sharp_5\ D\sharp_5\ B_4\ E_5$
11	Anchoring	$C_5\ D\sharp_5\ E_5\ C_5\ F\sharp_5\ G_5\ B_4\ D_5\ B_4\ G_4$
12	Run	$C_5\ D_5\ D\sharp_5\ E_5\ F_5\ F\sharp_5\ G_5\ D_5\ B_4\ G_5$
13	**Standard[c]**	$\mathbf{C_5\ E_5\ G_5\ E_5\ C_5\ G_5\ F\sharp_5\ D_5\ B_4\ C_5}$
14	Harmonic	$C_5\ E_5\ G_5\ E_5\ C_5\ G_5\ F_5\ D_5\ G_4\ C_5$
15	Anchoring	$C_5\ E_5\ G_5\ E_5\ C_5\ F\sharp_5\ G_5\ D_5\ B_4\ G_4$
16	Run	$C_5\ E_5\ G_5\ F\sharp_5\ F_5\ D_5\ B_4\ G_4\ B_4\ C_5$
17	**Standard[d]**	$\mathbf{B_4\ E_5\ G_5\ E_5\ B_4\ G_5\ F_5\ D_5\ A\sharp_4\ C\sharp_5}$
18	Harmonic	$B_4\ E_5\ G_5\ E_5\ B_4\ G_5\ F\sharp_5\ D\sharp_5\ B_4\ E_5$
19	Anchoring	$B_4\ E_5\ G_5\ F_5\ D_5\ A\sharp_4\ A_4\ C\sharp_5\ E_5\ D_5$
20	Run	$B_4\ E_5\ G_5\ F_5\ D_5\ A\sharp_4\ A_4\ B_4\ C\sharp_5\ D_5$

NOTE—Harmonic = sequences inducing a common chord progression; Anchoring = sequence contains nonchord tones that are captured by means of anchoring; Run = sequence contains nonchord tones that are captured in a run. (All chromatic notes are notated as sharps.)

[a]Derived from Cuddy et al. (1981), Expt. 2 S2,1
[b]Derived from Cuddy et al. (1981), Expt. 2 S4,1
[c]Derived from Cuddy et al. (1981), Expt. 1 # 13
[d]Derived from Cuddy et al. (1981), Expt. 1 # 24.

Stimulus Presentation

The sequences were presented in a different random order for each of the Nijmegen participants. The students from Northwestern University received only two random orders, because the experiment was run in two group sessions. Each sequence was played at a randomly determined pitch height varying between 6 semitones above and below their notated pitch in Table 1. The sequences were played through a Yamaha PSR-620 synthesizer using the Grand Piano sound. Both stimulus presentation and response collection were controlled by means of a program written in REALbasic running on a Macintosh G3 computer.

Procedure

The participants from the Nijmegen group performed the experiment individually in the following way: A participant was seated in front of a computer screen that displayed a window with buttons as shown in Figure 5. In the middle of the window, the stimulus is shown as a sequence divided into three groups of three tones followed by one single tone. Initially only the button in the corner right below, marked “start,” was enabled. When this button was pressed, its text changed into “next” and the “all” button was enabled. Upon

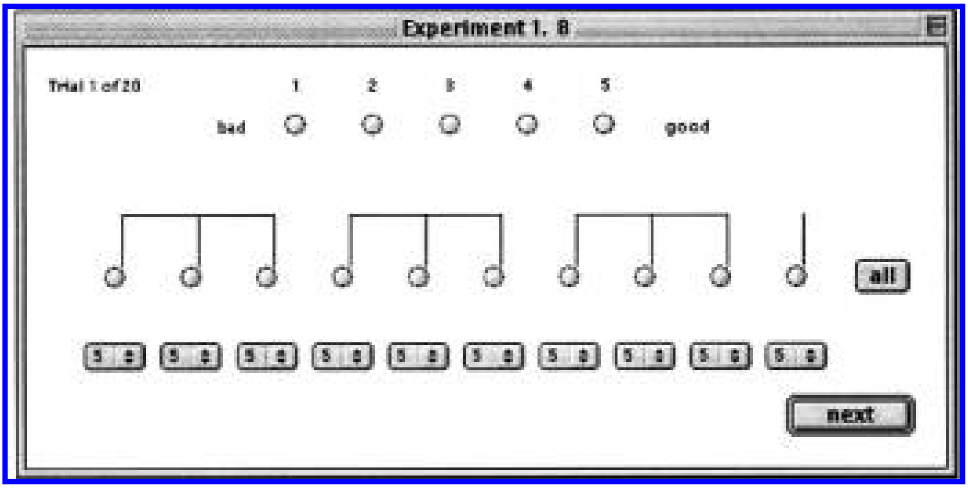

Fig. 5. Layout of the computer window used in the experiments. The window displays the temporal configuration of the tone sequence, a five-point rating scale for judging the global goodness of the sequence, and the pop-up menus below the tones, which serve as rating scales for the individual tones. In Experiment 2, the tone sequence contained one extra group of three tones.

pressing the "all" button, the sequence was presented. Next the five radio buttons at the top of the window, flanked with the texts "bad" on the left and "good" on the right, and the 10 pop-up menus below the notes were activated. The participant rated the global goodness or wellformedness of the sequence by activating one of the radio buttons on top of the screen. Next, in case the global rating was lower than 5, the subject indicated which tone or which tones sounded bad using the pop-up menus, which served as five-point rating scales for the individual tones. In doing these tasks, participants could listen to the stimulus as many times as desired. The Northwestern students heard each sequence twice and provided their ratings on response forms.

RESULTS

All participants enjoyed the experiment and experienced the tasks as interesting, musically relevant, and not particularly difficult. When asked what criteria they used to rate the sequence, participants gave answers like: "I tried to determine whether any of the tones were not in the key"; "I tried to determine whether or not the tones in the sequence were actually resolved"; "I paid attention to continuity in the phrase"; "Very out of place tones, dissonant leaps, unlikely continuations made the sequence sound less good." Participants mentioned that a sequence that sounded rather bad at first presentation often sounded better upon second hearing.

Some participants reported that a particularly bad tone (a "jarring" note) tends to influence the subsequent tones in the sense that they also appear to sound bad. More precisely, the listeners often found it difficult to tell whether the tones following a jarring tone were actually good or bad, that is, they could not tell whether these tones fit with the fragment that preceded the jarring tone. We shall come back to this in the Discussion.

Analysis of the Global Ratings

The mean global ratings for the stimuli are shown in Table 2. The mean global ratings for the four conditions (types of sequences), *standard, harmonic, anchoring,* and *run* averaged over all participants are 2.612, 4.238, 3.375, and 3.367, respectively. Figure 6 displays the results on the four conditions for the two groups of subjects. It indicates that although the Northwestern students' ratings for the last three conditions are somewhat lower, the overall pattern is the same: the standard sequences were rated lowest, the harmonic sequences highest, and the sequences in the anchoring and run conditions obtained a rating in between these two values. An analysis of variance (within-subject design) performed on the mean ratings of the four types of stimuli shows a significant effect for conditions, $F(3, 45) = 88.58$, $p < .0001$.

To determine the significance of the differences between the individual groups, planned comparisons were performed. T-tests, after Bonferroni adjustment (alpha = .008), show, for both groups of participants, significant differences for all six pairs of conditions, except for the difference between the anchoring and run condition.

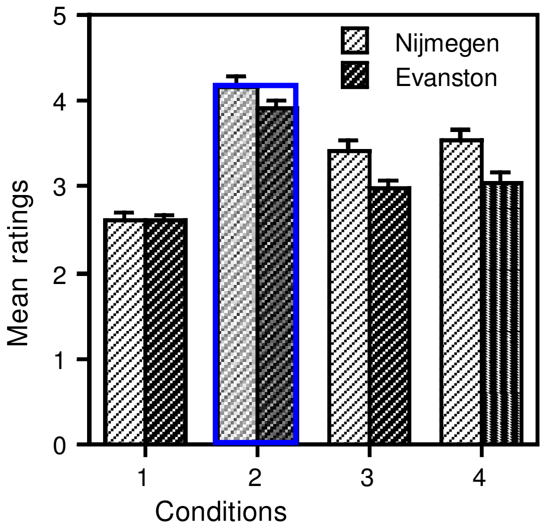

Fig. 6. Mean global ratings of the two groups of participants for the four categories of stimuli in Experiment 1: (1) Standard, (2) Harmonic, (3) Anchoring, and (4) Run.

Analysis of the Individual Tone Ratings

To compare the ratings of the individual tones with the ratings provided for the sequences as a whole (the global ratings), we calculated the mean tone rating per sequence by averaging the ratings over tones and subjects. The correlation between the mean tone ratings and the global ratings is .962 ($p < .001$), indicating that the judgment of the individual tones is well reflected in the global ratings.

Figure 7 shows the ratings of the individual notes for all 20 stimuli. The four rows represent the four conditions: standard, harmonic, anchoring, and run, each containing five sequences. A few typical characteristics of these response profiles should be mentioned. First, in several cases a tone that is rated low in the standard condition (first row) is also rated relatively low in the other conditions. Such is the case for the A in the first column, for the D♯ in the second column, for the F♯ in the fourth column, and the A♯ in the last column. Note, however, that in at least one case, the chromatic tone is rated just as high as the diatonic tones (Sequence 12). Second, note the differing ratings of the five sequences in the harmonic condition. The highest ratings are given to Sequences 6, 14, and 18 with an underlying I–V–I progression (18 is a minor mode), an intermediate rating to Sequence 2 with an underlying I–ii–V–I progression, and the lowest rating to Sequence 10, which induces a VI–V–i progression.

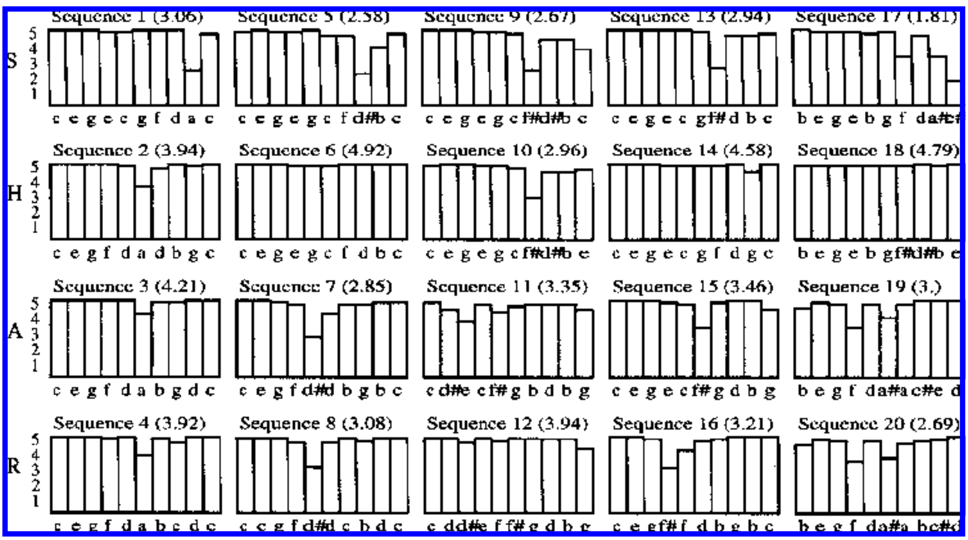

Fig. 7. Mean ratings of the individual tones in Experiment 1. All subjects (N = 48).

Comparison with Alternative Explanations

As mentioned in the Introduction, we have compared the explanation just advanced with three alternative explanations: one based on the tonal hierarchy concept of Krumhansl (1990), one on the formalization of part of the implication-realization model of Narmour (1990) by Schellenberg (1997), and one on the presence of specific features in the sequence.

To test how well the average position of the tones in the tonal hierarchy (stability) explains the ratings collected in the experiment, we have determined the stability of each separate tone in a sequence and calculated the mean of these stabilities, represented in the variable *MeanStab*. The tonal hierarchy used in the computations is that of the key of C major for Stimuli 1-16, E minor for Stimuli 17 and 18, and D major for Stimuli 19 and 20.

To test the predictions from Schellenberg's reduction of the implication–realization model of Narmour, we based ourselves on the two principles *pitch proximity* and *pitch reversal*, shown by Schellenberg (1997) to have the same explanatory power as the original principles, and have modified these variables such that they are applicable to an entire sequence in the following way:

1. To obtain a global measure of the pitch proximity principle, applicable to an entire melody, we have computed the mean of the absolute interval sizes, called *MeanInt*. According to the pitch proximity principle, the first interval should not be included, because the principle applies to two consecutive intervals, but this seems to be too dogmatic as the first interval will normally also be judged. Moreover, it would not make any difference for our stimuli because they all begin with the same two intervals.
2. To obtain a global measure of the *pitch reversal* principle, applicable to an entire melody, we have computed the pitch reversal implication for all successive pairs of intervals in a sequence using the quantification of the principle as displayed in Figure 4. Next we have calculated the mean pitch reversal implication by summing the individual values after adding + 1 to each value (to compensate for the fact that the local value may be –1, see Figure 4). As the tritone (interval of 6 semitones) is not dealt with in the principle, we have treated it in the same way as an interval of 7 semitones. This variable is called *PitchRev*.

To test the possible effect from specific features in the sequences we have determined three alternative variables: the number of contour changes, *ContChanges*; the average consonance value of the intervals in the sequence—based on the study of Malmberg (1918), as quantified by Krumhansl (1990, p. 57)—*ConsDis*, and the variation of the interval sizes in the sequence *VarInt* (defined as the standard deviation of the interval sizes). The values of the variables for the 20 stimuli is shown in Table 2.

To compare the predictive power of the alternative explanations with our explanation, we first calculated the Pearson correlation among the independent variables and the dependent variable (Table 3). Inspection of the correlation table reveals that only the variable Category is significantly correlated (r = -.69) with the dependent variable Rating and that several of the independent variables are highly intercorrelated. Although all other variables are not significantly correlated with Rating, we performed a hierarchical multiple regression in which the predictors associated with the four explanations are entered incrementally in the analysis (Tabachnik & Fidell, 2001, p. 133 ff.). Thus in the first step, the variable MeanStab, based on the tonal hierarchy notion, was entered; in the next step, the variables MeanInt and PitchRev, based on the implication-realization model, were added; in the third step, the variables ContChan, ConsDis, and VarInt were added; and in the final step the variable Category, related to our explanation, was added. The results of the analysis are shown in Table 4.

TABLE 2

Values of the Dependent and Independent Variables for the Stimuli Used in Experiment 1

Stim No.	Rating	Category	MeanStab	MeanInt	PitchRev	ContChanges	ConsDis	VarInt
1	3.06	4	4.94	3.78	1.5	4	5.59	1.4
2	3.94	1	4.51	3.78	1.75	4	5.32	1.03
3	4.21	2	4.50	3.56	1.5	5	6.24	1.57
4	3.92	3	4.62	2.67	1.375	3	7.8	1.15
5	2.58	4	4.75	3.56	1.687	6	6.12	1.64
6	4.92	1	4.86	3.56	2.062	6	6.01	1.57
7	2.85	2	4.31	2.67	1.375	2	7.64	1.15
8	3.08	3	4.49	2.22	1.562	3	8.71	0.92
9	2.67	4	4.59	3.78	1.812	6	6.37	1.69
10	2.96	1	4.40	4.22	2	6	5.46	1.4
11	3.35	2	4.15	3.67	1.375	5	7.38	2.11
12	3.94	3	3.99	2.56	1	2	9.12	2.31
13	2.94	4	4.71	3.33	1.5	4	6.49	1.7
14	4.58	1	5.09	4.22	1.625	4	5.09	1.69
15	3.46	2	4.59	3.67	1.25	3	6.27	1.33
16	3.21	3	4.33	2.67	1.375	2	7.51	1.25
17	1.81	4	3.60	4	1.5	4	5.91	1.7
18	4.79	1	3.67	4.11	1.5	4	5.72	1.85
19	3	2	4.02	3	1.375	3	6.94	1.15
20	2.69	3	4.04	2.56	1.375	2	8	1.26

NOTE—MeanStab = mean stability of the notes in the sequence; MeanInt = mean interval of the notes (proximity principle); PitchRev = expectancy based on the pitch reversal principle; ContChanges = number of contour changes; ConsDis = mean consonance of the intervals; VarInt = standard deviation of the interval sizes.

Table 3
Correlations Between the Dependent and Independent Variables of Experiment 1

	Global Ratings	Category	Mean Stab	Mean Int	Pitch Rev	Cons/ Dis	VarInt	Cont Changes
Global ratings	1.							
Category	-0.69**	1.						
MeanStab	0.25	0.01	1.					
MeanInt	0.16	-0.30	0.14	1.				
PitchRev	0.11	-0.28	0.41	0.52*	1.			
ConsDis	-0.17	0.31	-0.31	-0.91*	-0.62*	1.		
VarInt	0.18	-0.09	-0.21	0.35	0.18	-0.03	1.	
ContChanges	0.08	-0.10	0.34	0.70**	0.79**	-0.63**	0.29	1.

* $p < .05$; ** $p < .01$.

Table 4
Summary of a Hierarchical Regression Analysis on the Data of Experiment 1

Model	R^2	R^2 Adjusted	Standard Error	Change Statistics: R^2 Change	Change Statistics: F Change	p
1	.061	.008	.8080	.061	1.162 (1, 18)	.295
2	.082	-.090	.8471	.022	.189 (2, 16)	.830
3	.447	-.170	.8776	.117	.635 (3,13)	.605
4	.735	.580	.5256	.535	24.247 (1, 12)	.000

Note— Table shows the variance explained (R^2 change) by the three alternative explanations (see text) and by the harmonic model proposed in this study. The predictors added at each step are printed in bold. Model 1 Predictors: **MeanStab;** Model 2 Predictors: MeanStab, **MeanInt, PitchRev;** Model 3 Predictors: MeanStab, MeanInt, PitchRev, **ContChanges, ConsDis, VarInt;** Model 4 Predictors: MeanStab, MeanInt, PitchRev, ContChanges, ConsDis, Varint, **Category.**

The results of the regression analyses confirm the picture already presented by the correlations: the predictors related to the alternative explanations contribute only a very small and nonsignificant part to the explained variance, whereas the variable Category does explain a significant and substantive portion of the variance in the data, thereby corroborating the results of the analysis of variance reported earlier.

DISCUSSION

In this experiment, we have shown that "bad" sequences (sequences receiving low goodness ratings) can be transformed into "good" sequences, either by altering the alleged bad tones such that a harmonic analysis becomes feasible or by altering tones in the neighborhood of an alleged bad tone such that it can be anchored to a chord tone or incorporated in a run. These results support the basic idea that initiated this study, namely, that in order to perceive a tone sequence as a tonal melody, the listener must succeed in creating a harmonic analysis. If this analysis is hampered by the occurrence of nonchord tones, the sequence will receive a low rating. In addition, it was shown that the interference effect of nonchord tones is less if they are somehow assimilated by other tones in the sequence.

As regards the specific predictions made we may conclude that:

1. Sequences that allow a straightforward harmonic analysis yield a higher goodness rating than sequences that do not allow this (condition harmonic vs. the three other conditions). The data further suggest that a sequence is rated higher if the induced harmonic progression is more common.
2. Sequences containing nonchord tones that are somehow assimilated, either by means of the mechanism of anchoring or by being assimilated in a run, are rated higher than sequences that do not accommodate nonchord tones in these ways (condition standard vs. conditions anchoring and run). In the anchoring and run conditions, the nonchord tones are rated lower than the chord tones. This important finding indicates that, in general, nonchord tones, even when anchored or being part of a run, are still experienced as having some interference effect. The results also indicate that anchoring is not an all or none process but that its operation depends on the actual tones involved. For instance, Sequence 12, in which the F♯ is followed by a G, is rated higher than Sequence 16, in which the F♯ is followed by F.

In addition, we may conclude that the goodness rating of the individual tones is a valuable response indicator because it provides a fairly detailed picture of which tones are difficult to incorporate in a musical interpretation.

Two other observations, made while conducting this study, are worth mentioning. First, the finding that a series often sounds better when heard for the second time may indicate that the process of perception is a relatively slow and dynamic process in which the listener tries to make sense of the series of tones by developing different hypotheses about how to incorporate the tones in a coherent percept. Second, the individual tone ratings shown in Figure 7 indicate that often not only the jarring tone but also the subsequent tones receive a lower rating. One participant expressed this very clearly: “I don’t seem to be able to interpret the tones that follow the bad tone; I cannot decide whether or not they fit in the sequence as a whole.” This result suggests that it is not possible to relate tones to previous tones if the harmonic analysis is interrupted and demonstrates the critical significance of an effective harmonic interpretation in music perception.

A discussion of the comparison between the predictions of our model with predictions based on the other explanations is postponed until we report a similar analysis on the data of the next experiment.

Because in this experiment only a limited number of sequences was used, and because the potential variables were not varied systematically, it is difficult to tell to what extent these findings can be generalized to other sequences. Besides, we would like to establish whether there exist alternative mechanisms that reduce the interference effect of nonchord tones. These issues are further examined in Experiment 2.

Experiment 2

In the previous experiment, we have provided some evidence that goodness ratings of tone sequences depend on whether the listener is able to create a harmonic analysis, and that the positioning and context of nonchord tones are critical in determining whether such an analysis will succeed. In this experiment, we examine the conditions that play a role in the induction of an uninterrupted harmonic analysis in a more systematic way by using an extended set of stimuli in which all possible nonchord tones are introduced. For this purpose, we started with a prototypical sequence based on a I–V–I progression from which a number of new sequences were derived by replacing one tone in the V section by another tone within the range of an octave. With this set of stimuli, we want to verify the findings of Experiment 1 and to discover whether there are additional factors that interact with the factors discerned so far.

METHOD

Subjects

Twenty-one adult listeners, all from the Nijmegen university community, participated in the experiment (median age = 27 years). All participants were at least moderately trained in music, with minimally 5 years of lessons on a musical instrument or in singing (mean = 11.1 years, *SD* = 5.9 years).

Construction of the Stimuli

Beginning with the prototypical sequence C_5 E_5 C_5 G_5 E_5 C_5 B_4 D_5 G_5 F_5 D_5 B_4 C_5, based on a I–V–I progression, we constructed a set of sequences by substituting the tone F_5 for each of the tones between G_4 and A_5 (a fifth below and above the tone following F_5). From this set, we eliminated the sequences that contained two consecutive identical tones (the immediate repetition of a tone does not introduce any extra harmonic factor, whereas it was found in a pilot experiment that a sequence containing a repeated tone is very conspicuous and difficult to judge in relation to sequences not containing a repetition). In a similar fashion, the last D_5 and the last B_4 of the sequence were substituted, thus obtaining the set of 35 stimuli displayed in Table 5. In the set, all possible nonchord tones were introduced at three different positions in the V segment of the sequences.

To control the metrical interpretation, the stimuli were generated with timing and dynamic features obtained from an actual performance of the prototypical stimulus played in a 12/8 meter, that is, very similar to the metrical configuration of Experiment 1, except that this sequence has four instead of three groups of three tones. The average interonset interval between the tones of the metrical template was 600 ms, whereas the velocity (intensity) of

Table 5
Stimuli Used in Experiment 2

No.	Sequence					Mean Rating	Rating Subst.
1		$G\sharp_4$				2.19	2.
2		A_4				3.24	3.33
3		$A\sharp_4$				3.05	2.81
4		B_4				4.43	4.57
5		C_5				3.67	3.76
6		$C\sharp_5$				2.33	2.29
7	C_5 E_5 C_5 G_5 E_5 C_5 B_4 D_5 G_5	$D\sharp_5$	D_5	B_4	C_5	2.67	2.48
8		E_5				3.38	3.38
9		F_5				5.	5.
10		$F\sharp_5$				2.67	2.71
11		$G\sharp_5$				2.81	2.57
12		A_5				3.76	3.95
13			E_4			1.9	1.81
14			$F\sharp_4$			1.57	1.57
15			G_4			4.67	4.76
16			$G\sharp_4$			2.43	2.29
17			A_4			4.57	4.67
18	C_5 E_5 C_5 G_5 E_5 C_5 B_4 D_5 G_5	F_5	$A\sharp_4$	B_4	C_5	3.	3.1
19			C_5			3.33	3.48
20			$C\sharp_5$			2.33	2.05
21			$D\sharp_5$			2.57	2.38
22			E_5			4.	4.24
23			$F\sharp_5$			2.57	2.14
24				F_4		2.48	3.1
25				$F\sharp_4$		1.86	1.71
26				G_4		4.52	4.48
27				$G\sharp_4$		1.81	1.76
28				A_4		2.95	2.95
29	C_5 E_5 C_5 G_5 E_5 C_5 B_4 D_5 G_5	F_5	D_5	$A\sharp_4$	C_5	2.33	2.19
30				$C\sharp_5$		2.48	2.19
31				$D\sharp_5$		3.19	2.81
32				E_5		4.67	4.9
33				F_5		3.86	3.67
34				$F\sharp_5$		2.05	1.71
35				G_5		4.62	4.86

Note—To clarify the construction of the stimuli, the complete sequence is shown only three times, once for each position in which tones are substituted. Penultimate column: mean global rating. Last column: mean rating of substituted tone. Stimulus 9 is the prototypical sequence from which all other sequences are derived.

the first tones of the groups (downbeats) was approximately 60 (on a scale from 1 to 127), that of the second tone 38, and that of the last tone 33. All stimuli were generated with this timing and velocity pattern.

Stimulus Presentation and Procedure

Stimulus presentation and procedure were the same as in Experiment 1, except that the window on the computer was adapted for the longer experimental stimuli and the subject's task was to judge whether the sequence might occur in a simple piece of tonal music. As in the previous experiment, participants indicated which tone(s) did not seem to fit well in the sequence by using the five-point rating scales below each tone. To get familiarized with the procedure, the participants practiced with six stimuli not used in the experiment.

RESULTS

The participants liked the musical relevance of the task, which they found relatively easy to perform. Some participants remarked that sometimes it was not so much the substituted tone that was wrong but rather the way the sequence continued after that tone.

Between-subject reliability is quite high, as witnessed by a split-half correlation of .94. Correlation between the mean global ratings and the mean of the individual tone ratings is very high: .973, as in the preceding experiment. Also the correlation between the global ratings and the ratings of the substituted tone is very high: .984, indicating that the quality of the sequences is mainly determined by the substituted tone.

The mean global ratings of the stimuli and the rating of the substituted tone are shown in Table 5. In Table 6, the stimuli are annotated and ordered according to rating.

Inspection of Table 6 shows that sequences containing only chord tones are rated highest, those containing a local nonchord tone are rated lower, and those containing a nonkey tone are rated lowest. More importantly, the table also shows that sequences containing a nonchord tone (either diatonic or chromatic), which can be accommodated either in a harmonic interpretation or by means of anchoring, are rated higher than those that cannot be captured by one of these mechanisms. To make this apparent, in Column 3 of Table 6 we have indicated whether a nonchord tone can be accommodated in an alternative, less likely, harmonic analysis (denoted with the abbreviation AltHarm) or may be resolved by means of anchoring (denoted Anch). Next we have distinguished three conditions by dividing the sequences into three categories, labeled 1, 2, and 3 in the last column: (1) sequences that induce a I–V–I progression (6 sequences); (2) sequences in which a nonchord tone is captured either by an alternative harmonization or by anchoring (19 sequences); (3) sequences containing a nonchord tone that cannot be captured by either of the two mechanisms (10 sequences). The mean ratings for the stimuli in these three conditions are respectively

Table 6

Classification and Ratings of the Stimuli in Experiment 2 Ordered by Rating

No.	Sequence	Type	Glob	Subst	C
9	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ B_4\ C_5$	H (prototypical sequence)	5.	5.	1
32	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{E}_5\ C_5$	H	4.67	4.76	1
15	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ G_4\ B_4\ C_5$	H	4.67	4.9	1
35	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ G_5\ C_5$	H	4.62	4.86	1
17	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{A}_4\ B_4\ C_5$	AltHarm IV-V-I/Run	4.57	4.67	2
26	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ G_4\ C_5$	H	4.52	4.48	1
4	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ B_4\ D_5\ B_4\ C_5$	H	4.43	4.57	1
22	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\text{-}\mathbf{E}_5\ B_4\ C_5$	Anch	4.	4.24	2
33	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ F_5\ C_5$	AltHarm ii-I	3.86	3.67	2
12	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{A}_5\ D_5\ B_4\ C_5$	AltHarm ii-V-I	3.76	3.95	2
5	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{C}_5\text{-}D_5\ B_4\ C_5$	Anch	3.67	3.76	2
8	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{E}_5\text{-}D_5\ B_4\ C_5$	Anch	3.38	3.38	2
19	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{C}_5\text{-}B_4\ C_5$	Anch/AltHarm IV-V-I	3.33	3.48	2
2	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{A}_4\text{-}D_5\text{-}B_4\ C_5$	DelAnch/Alt Harm ii-V	3.24	3.33	2
31	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{D\sharp}_5\ C_5$	AltHarm: V-i (minor)	3.19	2.81	2
3	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{A\sharp}_4\text{-}D_5\text{-}B_4\ C_5$	DelAnch/Run	3.05	2.81	2
18	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{A\sharp}_4\text{-}B_4\ C_5$	Anch	3.	3.1	2
28	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{A}_4\ C_5$	AltHarm ii-I	2.95	2.95	2
11	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{G\sharp}_5\ D_5\ B_4\ C_5$	AltHarm III-I	2.81	2.57	2
10	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{F\sharp}_5\ D_5\ B_4\ C_5$	-	2.67	2.48	3
7	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{D\sharp}_5\text{-}D_5\ B_4\ C_5$	Anch	2.67	2.71	2
23	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{F\sharp}_5\ B_4\ C_5$	-	2.57	2.38	3
21	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{D\sharp}_5\ B_4\ C_5$	AltHarm V-I (minor)	2.57	2.14	2
30	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{C\sharp}_5\ C_5$	AltHarm ii-I/ Run	2.48	3.1	2
24	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ F_4\ C_5$	AltHarm ii-I;	2.48	2.19	2
16	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{G\sharp}_4\ B_4\ C_5$	-	2.43	2.29	3
29	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{A\sharp}_4\ C_5$	-	2.33	2.29	3
20	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{C\sharp}_5\text{-}B_4\text{-}C_5$	Anch/DelAnch (C#-C)	2.33	2.05	2
6	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{C\sharp}_5\text{-}D_5\ B_4\ C_5$	Anch	2.33	2.19	2
1	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ \mathbf{G\sharp}_4\ D_5\ B_4\ C_5$	-	2.19	2.	3
34	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{F\sharp}_5\ C_5$	-	2.05	1.71	3
13	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{E}_4\ B_4\ C_5$	-	1.9	1.81	3
25	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{F\sharp}_4\ C_5$	-	1.86	1.71	3
27	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ D_5\ \mathbf{G\sharp}_4\ C_5$	-	1.81	1.76	3
14	$C_5\ E_5\ C_5\ G_5\ E_5\ C_5\ B_4\ D_5\ G_5\ F_5\ \mathbf{F\sharp}_4\ B_4\ C_5$	-	1.57	1.57	3

Note—Glob = global rating; Subst = rating of the substituted tone; H = harmonic; Anch = Anchoring (indicated with a dash between the tones); DelAnch = delayed anchoring; AltHarm = alternative harmonic analysis; - = does not allow an harmonic analysis; C = condition.

4.65, 3.14, and 2.14 (Figure 8). An analysis of variance performed on the subjects' mean ratings of the three conditions exhibits a highly significant effect for conditions, $F(2, 19) = 91.10$, $p < .0001$. Planned comparisons, applying Bonferroni's adjustment (alpha = .016), indicate that all three pairwise comparisons are significantly different (all p values < .0001).

To examine whether the position of the substituted tone affected the ratings, we performed a two-way analysis of variance with Condition (3 levels) and Position (3 levels; first, second, or third tone in penultimate beat) as within factors. This analysis shows a significant effect for Condition (as in the former analysis), whereas the effect of Position and the interaction between the factors are not significant.

Comparison with Alternative Explanations

Similar to the first experiment, we have compared our explanation of the data with the three alternative explanations. Table 7 shows the values of the dependent and independent variables for the 35 stimuli, and Table 8 presents the intercorrelations between the independent and dependent variables.

It appears that two variables correlate significantly with the dependent variable rating, namely Category (r = -.85), and MeanStab (r = .528), indicating that these factors on their own (without taking into account correla-

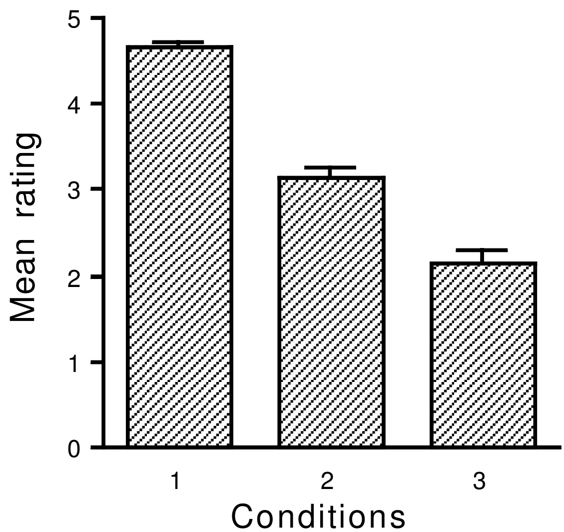

Fig. 8. Mean global ratings for the three conditions in Experiment 2. Condition 1 contains sequences clearly inducing a I–V–I progression; Condition 2 contains sequences in which a nonchord tone is assimilated by anchoring or by an alternative harmonization; Condition 3 contains sequences containing a nonchord tone that cannot be captured. See text for explanation.

TABLE 7
Values of the Dependent and Independent Variables for the Stimuli Used in Experiment 2

StNo.	Mean Rating	Category	Mean Stab	Mean Int	Pitch Rev	Cont Chan	Cons Dis	Var Int
1	2.19	3	4.59	4.33	1.68	8	6.69	2.62
2	3.24	2	4.69	4.17	1.73	8	6.12	2.37
3	3.05	2	4.58	4.	1.73	8	5.92	2.16
4	4.43	1	4.63	3.83	1.73	8	6.21	1.99
5	3.67	2	4.89	3.67	1.86	8	6.09	1.89
6	2.33	2	4.58	3.5	1.86	8	6.75	1.85
7	2.67	2	4.58	3.33	1.64	6	6.45	1.7
8	3.38	2	4.74	3.33	1.64	6	6.47	1.6
9	5.	1	4.72	3.33	1.5	6	6.47	1.6
10	2.67	3	4.6	3.33	1.5	6	6.45	1.7
11	2.81	2	4.59	3.5	1.41	6	6.75	1.85
12	3.76	2	4.69	3.67	1.41	6	6.09	1.89
13	1.9	3	4.79	4.5	1.36	6	6.48	3.18
14	1.57	3	4.65	4.17	1.45	6	6.5	2.64
15	4.67	1	4.85	4.	1.45	6	6.49	2.42
16	2.43	3	4.64	3.83	1.45	6	6.32	2.23
17	4.57	2	4.73	3.67	1.45	6	6.69	2.09
18	3.	2	4.63	3.5	1.45	6	6.48	2.02
19	3.33	2	4.94	3.33	1.5	6	6.56	1.8
20	2.33	2	4.62	3.33	1.5	6	6.54	1.65
21	2.57	2	4.63	3.33	1.36	6	6.54	1.65
22	4.	2	4.79	3.33	1.36	6	6.56	1.8
23	2.57	3	4.65	3.5	1.59	8	6.48	2.02
24	2.48	2	4.81	4.33	1.59	6	5.55	2.21
25	1.86	3	4.69	4.17	1.59	6	6.2	1.95
26	4.52	1	4.9	4.	1.59	6	5.4	1.73
27	1.81	3	4.68	3.83	1.45	6	6.05	1.57
28	2.95	2	4.78	3.67	1.5	6	5.78	1.49
29	2.33	3	4.67	3.5	1.5	6	6.15	1.5
30	2.48	2	4.67	3.17	1.36	5	6.86	1.72
31	3.19	2	4.68	3.33	1.64	7	6.46	1.6
32	4.67	1	4.83	3.5	1.64	7	6.15	1.5
33	3.86	2	4.81	3.67	1.64	7	5.78	1.49
34	2.05	3	4.69	3.83	1.64	7	6.05	1.57
35	4.62	1	4.9	4.	1.64	7	5.4	1.73

NOTE—MeanStab = mean stability of the notes in the sequence; MeanInt = mean interval of the notes (proximity principle); PitchRev = expectancy based on the pitch reversal principle; ContChanges = number of contour changes; ConsDis = mean consonance of the intervals; VarInt = standard deviation of the interval sizes.

tions with other variables) respectively explain 72.2%, and 27.9% of the variance in the data. The correlations of the other variables with the dependent variable are very low and statistically not significant. To assess the relative contribution of the variables Category and MeanStab, we performed two Hierarchical Multiple Regressions, one in which the variable MeanStab

TABLE 8
Correlations Between the Dependent and Independent Variables of Experiment 2

	Rating	Category	Mean Stab	Mean Int	Pitch Rev	Cont Changes	ConsDis	Var Int
Rating	1.							
Category	-0.85**	1.						
MeanStab	0.53**	-0.43**	1.					
Mean Int	-0.18	0.19	0.16	1.				
PitchRev	0.12	-0.17	0.01	0.17	1.			
ContChanges	0.10	-0.05	-0.12	0.25	0.82**	1.		
ConsDis	-0.24	0.25	-0.49**	-0.42*	-0.30	-0.15	1.	
VarInt	-0.24	0.27	-0.06	0.72**	-0.08	0.15	0.20	1.

* $p < .05$; ** $p < .01$

was entered in the first step and the variable Category in the second step, and another one using the reverse entry order. The results of these analyses are shown in Table 9.

The analyses reveal that the variable Category adds 47.7% to the variance explained (R^2 change), if the variable MeanStab is entered first (Analysis 1), whereas the variable MeanStab adds only 3.3% to the explained variance if the variable Category is entered first (Analysis 2). From this it may be concluded that although MeanStab does explain part of the data, virtually all of it is also explained by Category. Apparently, the variables MeanStab and Category (or better the underlying models), both describe aspects of the data that are relevant in explaining the responses.

TABLE 9
Summary of Two Hierarchical Regression Analyses Performed on the Data of Experiment 2

ANALYSIS 1			Standard	Change Statistics		
Model	R^2	R^2 Adjusted	Error	R^2 Change	F Change	p
1	.280	.258	.8365	.280	12.816 (1, 33)	.001
2	.756	.741	.4943	.476	62.492 (1, 32)	.000

ANALYSIS 2			Standard	Change Statistics		
Model	R^2	R^2 Adjusted	Error	R^2 Change	F Change	p
1	.723	.715	.5186	.723	86.211 (1, 33)	.000
2	.756	.741	.4943	.033	4.315 (1, 32)	.046

NOTE—In Analysis 1, the variable MeanStab is entered in the first step and the variable Category added in the second. In Analysis 2, the ordering of entering is reversed.

DISCUSSION

In this experiment, we used an extended set of stimuli to verify the hypothesis that to perceive a sequence of tones as a melody the listener must succeed in making a harmonic analysis. If we compare the mean rating of the stimuli in Condition 1, containing sequences that strongly induce a I-V-I progression (4.65), with that of the sequences in Condition 3 for which the formation of a complete progression seems impossible (2.14), we may conclude that the data support this hypothesis.

The second purpose of the experiment was to discover the conditions that enable a harmonic analysis in spite of the occurrence of a nonchord tone. It was hypothesized that listeners either attempt to make an alternative harmonic analysis that transforms the nonchord tone into a chord tone, or try to anchor the nonchord tone to a subsequent chord tone. To different extents, these conditions are fulfilled in the sequences in Condition 2, which on average were indeed rated higher than the sequences in Condition 3 (mean rating 3.14 vs. 2.14). Thus the data indicate that the degree to which nonchord tones interfere with the formation of a complete harmonic analysis strongly depends on the context in which these nonchord tones appear. This context is rather local when the mechanism of anchoring is applied, but more global in cases where the nonchord tone is transformed into a chord tone by assigning an alternative harmonic analysis. Although the sequences in Condition 2 are rated higher than those in Condition 3, they are nevertheless rated lower than the sequences in Condition 1. This may be because in Condition 2 a harmonic analysis cannot be done as readily as in Condition 1, requiring additional processing to arrive at a musical interpretation.

The sequences in Condition 2 represent a rather extensive class of stimuli that are supposed to be processed in different ways to arrive at a harmonic analysis: either by means of anchoring or by means of developing a less common harmonic analysis. As shown in Table 6, for some of the stimuli, there is theoretically more than one solution to the problem, and it is not clear which of those solutions is actually applied by the listeners. It is conceivable that these sequences are processed in both ways and that their results are maintained as alternative codes, reflecting the basic ambiguity of musical interpretation. Anyway, it is clear that the details of this processing are still not well understood and must be further examined. As regards the mechanism of anchoring, the data imply that it is not an all-or-none process: chromatic tones are generally less well anchored than diatonic nonchord tones. With respect to the suggested alternative harmonic analyses, the data indicate that the less common the harmonic progression, the lower the rating; compare for instance stimuli 17 (I–IV–V–I, rating 4.57) and 24 (I–V–ii–I, rating 2.48). As we don't have sufficient data to examine this properly, we cannot go into this any further.

The last goal of this experiment was to examine whether there are other conditions besides anchoring that may facilitate the incorporation of a nonchord tone in a harmonic analysis. Mainly for that purpose, we introduced all tones within an octave in the stimuli. On the basis of the analysis of the stimuli in Table 6 and the successful prediction of the responses based on that analysis, we believe that there are no other alternative mechanisms that enable the assimilation of nonchord tones.

The comparison between the explanation of the data based on the mechanisms proposed in this study and those based on some alternative explanations, indicates that (a) the variable MeanStab, associated with Krumhansl's tonal hierarchy concept, was found to explain part of the variance, but the factor did not explain any variance not already explained by our model. These findings are understandable as MeanStab is correlated with Category ($r = .43$) and thus apparently measures an attribute that is related to harmonic interpretation. Indeed, as MeanStab is related to the average position of the tones in the key hierarchy, it will be higher if the sequence does not contain a nonkey tone and lower if it does (especially since most of the nonchord tones in the stimulus set are nondiatonic tones); (b) the predictions based on the principles derived from Narmour's implication-realization model do not come out right; (c) the variables number of contour changes, average consonance, and variability of the interval sizes do not contribute to explaining the variance in the data. We return to this issue later.

General Discussion

In two experiments, listeners judged the melodic goodness of tone sequences differing in the extent in which they, at least theoretically, induce a harmonic interpretation. The experiments yielded two results: (1) they revealed the factors that play a role in making a harmonic interpretation; (2) they provided experimental evidence that the goodness ratings can be explained by the effort needed to arrive at a harmonic analysis, as well as by the type of harmonic analysis effectively generated, as specified later.

Two global classes of single (unaccompanied) tone sequences were used in the study: (a) sequences whose constituents can be conceived as a succession of arpeggiated chords, that is, once a harmonic interpretation has been made, all tones are elements of one of the chords, and (b) sequences for which (a) does not hold; that is, after a harmonic interpretation has been made, there are tones left that are not part of one of the chords, usually called nonchord tones. Comparing the goodness judgments of these two classes of sequences, it is found that the former receive a much higher rating than the latter. Furthermore, ratings of more common progressions of chords (e.g., I–V–I) are higher than those of less common chord progressions (e.g., I–ii–I).

Next, those tone sequences containing nonchord tones were divided into three subclasses: (a) sequences containing nonchord tones that can be anchored to a subsequent chord tone, (b) sequences in which the nonchord tones are part of a run, (c) sequences for which the conditions under (a) and (b) do not apply. The ratings of the sequences in subclasses (a) and (b) are significantly higher than those for the sequences in subclass (c). Ratings of the sequences in classes (a) and (b) did not differ. It is interesting that the sequences for which the nonchord tones can be anchored or are part of a run were, in contrast to what we had expected, not rated as high as those that do not contain nonchord tones. These lower ratings may reflect the additional effort needed to assimilate these tones in the development of a harmonic interpretation.

The results of this study can be summarized as follows: (a) the formation of a harmonic interpretation is of paramount importance in the processing of tonal music, and largely determines the goodness ratings of these sequences, (b) nonchord tones may seriously hamper the development of a harmonic analysis, (c) the interference effect of nonchord tones depends on their context, and is lessened if the nonchord tones can be assimilated either by means of anchoring or by being conceived as part of a run. These results corroborate those of Cuddy et al. (1981) by showing that perceptual responses to tone sequences are greatly determined by the tonal structure of the sequences. In addition, they reveal the conditions in which nonchord tones are "assimilated" to chord tones and thus do not hamper the creation of a harmonic interpretation.

The incremental processes that give rise to a harmonic interpretation may now tentatively be described as follows. After having heard the first few tones of the sequence, the listener attempts to detect the underlying chord. This chord is then assumed to be a tonic chord, thus yielding a preliminary key candidate. With the (provisional) key is associated a whole set of expectations pertaining to the occurrence of tones (e.g., diatonic tones are more likely than chromatic tones) and chords (chords on the first, fourth, and fifth scale degree are more likely than chords on the other degrees), and so on. These expectations also include expectations regarding the order of tones and chords. For instance, if the listener has recognized a series of tones as a dominant seventh chord, an expectation for elements of the tonic chord, or less likely for elements of the chord on the sixth degree, will be created. If these expectations are fulfilled, the creation of a harmonic interpretation will succeed and the tone sequence will be perceived as an acceptable musical event.

As we have seen, the main difficulty in developing a harmonic interpretation is the occurrence of nonchord tones: tones that do not fit in the presently activated chord. If a nonchord tone appears, the listener will search for a solution that allows the incorporation of the nonchord tone in the present harmonic frame, notably by establishing whether the nonchord tone is resolved, by anchoring to a subsequent chord tone, or incorporated in a run. Thus it is essential that the nonchord tone is somehow incorporated in the harmonic frame. If this does not happen, the formation of a valid harmonic interpretation fails and the sequence will receive a low rating. Since the number of possible continuations tends to be highly constrained, some nonchord tones (especially chromatic tones) may appear to be completely incompatible with the present harmonic analysis and therefore be heard as a wrong tone.

In a comparison of our explanation with a few alternative ones, it was found that only the explanation based on Krumhansl's tonal hierarchy concept could partly account for the data collected in Experiment 2. In the Discussion of that experiment we have argued that this is probably due to the fact that the average stability of the tones in a sequence is related to the harmonic factors operative in our explanation.

Specific features in the stimuli such as the number of contour changes, the average consonance of the intervals, and the variation of the interval sizes appeared not to contribute to explaining the data.

The finding that the data cannot be explained by the two principles that according to Schellenberg (1997) represent an aspect of the implication-realization theory of Narmour (1990, 1992) may in first instance be surprising given the considerable body of research showing the importance of these principles in predicting expectancies listeners create when hearing an implicative interval. However, this result should be evaluated in the light of the following considerations. In the studies testing the innate aspects of Narmour's model, the principles were tested and shown to be operative in predicting the expectations concerning the last tone of three-tone sequences (Cuddy & Lunney, 1995) or of the last tone of initial fragments of actual melodies (Krumhansl, 1995; Schellenberg, 1996, 1997). As mentioned earlier, in the latter studies the influence of metrical and harmonic factors was held constant as much as possible, so that the role of these factors was minimized. In contrast, in the present research the main variation concerns the position and context of nonchord tones, whereas factors related to the contour were not systematically varied. Moreover, in the studies testing Narmour's principles, listeners focused on the final three tones when rating how well the last tone continued the melodic fragment, whereas in our experiments, which tested a model of melody processing, listeners rated the goodness of entire melodies. For these reasons, it is understandable that we did not find significant effects of Narmour's principles in the present research. Furthermore, it should be noted that the expectations based on Narmour's principles pertain to contour constraints that are less restrictive (e.g., listeners have a general expectation for a small interval, or after having heard a large ascending interval expect a subsequent smaller descending interval) than those based on our model in which listeners' expectations are supposed to depend on the previous chord and/or the occurrence of a nonchord tone implying a specific resolution. We want to emphasize that this test of Schellenberg's principles only bears on a relatively small aspect of Narmour's comprehensive implication-realization theory and that the results have no implication for this theory as a whole.

In conclusion then, our results seem to indicate that in the processing of tonal melodies meter and harmony are the primary frames in which a tonal melody is described, and that the precise contour of the melody may only be coded in some later stage if required. Usually listeners can tell whether a melody is a common tonal melody before they can reproduce it. But even if participants are explicitly asked to reproduce a melody exactly, it appears that they have trouble doing that and instead provide a paraphrase compatible with its metrical and harmonic frame (Sloboda & Parker, 1985).

Although the results reported in this study give an impression of the mechanisms that play a role in the induction of harmony, additional research is needed to explicate the conditions in which anchoring and incorporation in a run are effective in assimilating nonchord tones in the generation of a harmonic analysis. Using the insights gained so far, we are presently developing a computational model that describes how a harmonic frame is developed while a sequence evolves, which expectations are generated at different points in the sequence, and which tones will lead to a breakdown of the harmonic analysis.[2]

References

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16,* 485–518.

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. In P. M. Todd & D. G. Loy (Eds.), *Music and connectionism.* Cambridge, MA: MIT Press.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception, 13,* 383–401.

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

2. We thank Ric Ashley for his advice in constructing the stimuli and for allowing us to run the experiment in his music cognition classes at Northwestern University, Glenn Schellenberg for his advice and help in defining the factors related to Narmour's model, and Eugene Narmour for several thoughtful remarks pertaining to the interpretation of the results. We gratefully acknowledge the support and encouragement of Bob Gjerdingen in conducting the comparison studies. We thank Jules Ellis for his advice on the statistical analyses, Hubert Voogd for helping develop the program that ran the experiments. and René van Egmond and Hank Heijink for a critical reading of an earlier version of this report.

Cuddy, L. L. , & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics, 57*, 451–462.

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

Gjerdingen, R. O. (1990). Categorization of musical patterns by self-organizing neuron-like networks. *Music Perception, 8,* 339–370.

Holleran, S., Jones, M. R., & Butler, D. (1995). Perceiving implied harmony: The influence of melodic and harmonic context. *Journal of Experimental Psychology: LMC, 21,* 737–753.

Jansen, E. L., & Povel, D. J. (1999). Mechanisms in the perception of accented melodic sequences. *Proceedings of the 1999 Conference of the Society for Music Perception and Cognition* (p. 29). Evanston, IL: Society for Music Perception and Cognition.

Jansen E. L., & Povel, D. J. (2000). The role of implied harmony in the perception of brief tone sequences. In C. Woods, G. Luck, R. Brochard, F. Seddon, & J. A. Sloboda (Eds.) *Proceedings of the Sixth International Conference on Music Perception and Cognition* [CD-ROM] (pp. 927–934). Keele, Staffordshire, UK: Department of Psychology.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum, 17*, 53–80.

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.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception, 5,* 315–350.

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

Longuet-Higgins, H. C., & Steedman, M. J. (1971). On interpreting Bach. In B. Meltzer, & D. Michie (Eds.), *Machine intelligence* Edinburgh: University Press.

Malmberg, C. F. (1918). The perception of consonance and dissonance. *Psychological Monographs*, *25*, (2, whole No. 108), 93–133.

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

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

Piston, W., & Devoto, M. (1989). *Harmony*. London: Victor Gollancz. (Original work published 1941)

Platt, J. R., & Racine, R. J. (1994). Detection of implied harmony changes in triadic melodies. *Music Perception, 11,* 243–264.

Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology Human Perception and Performance*, *7*, 3–18.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411–440.

Povel, D. J., & Egmond, R. van (1993). The function of accompanying chords in the recognition of musical fragments. *Music Perception, 11,* 101–115

Povel, D. J., & Jansen, E. (2000). Towards an on-line model of music perception. In C. Woods, G. Luck, R. Brochard, F. Seddon, & J. A. Sloboda (Eds.) *Proceedings of the Sixth International Conference on Music Perception and Cognition.* Keele, UK: Keele University Department of Psychology.

Povel D. J., & Jansen, E. (2001). Perceptual mechanisms in music perception. *Music Perception, 19,* 169–198.

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.

Sloboda, J. A., & Parker, D. H. H. (1985). Immediate recall of melodies. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition.* London: Academic Press.

Tabachnik, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* Boston: Allyn & Bacon.

Temperley D. (1997). An algorithm for harmonic analysis. *Music Perception, 15*, 31–68.

Thompson, W. F. (1993). Modeling perceived relationships between melody, harmony, and key. *Perception & Psychophysics, 53,* 13–24.

Thompson, W. F., & Cuddy, L. L. (1989). Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony. *Music Perception, 7,* 151–168.

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.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107,* 885–913.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants and adults sensitivity to Western musical structure. *Journal of Experimental Psychology, Human Perception & Performance, 18*, 394–402.

Trehub, S. E., Thorpe, L. A., & Trainor, L. J. (1990). Infant's perception of good and bad melodies. *Psychomusicology, 9*, 5–19.

Van Dyke Bingham, W. (1910). Studies in melody. *Psychological Review*, Monograph Supplements. Vol. XII, Whole No. 50.