

INFANTS' PERCEPTION OF CONSONANCE AND DISSONANCE IN MUSIC

Marcel R. Zentner¹

Jerome Kagan

Harvard University

The origins of the perception of consonance and dissonance in music are a matter of debate. The present study examined the hypothesis of an innate preferential bias favoring consonance over dissonance by exposing 4 month old infants to consonant and dissonant versions of two melodies. Infants looked significantly longer at the source of sound and were less motorically active to consonant compared with dissonant versions of each melody. Further, fretting and turning away from the music source occurred more frequently during the dissonant than the consonant versions. The results suggest that infants are biologically prepared to treat consonance as perceptually more pleasing than dissonance.

Infants Consonance Dissonance Music Perception

INTRODUCTION

The possibility of perceptual universals in music has been a question of interest for a long time. After many centuries of speculative debate this question is now being empirically addressed. The study of infants represents one way to explore musical universals. This research investigates the hypothesis of an innate bias for consonance over dissonance.

Consonance and dissonance refer to the subjective judgement of a listener, exposed to

a combination of *two or more* frequencies occurring simultaneously. When the combination is experienced as pleasant, the sound is classified as consonant. When the combination of frequencies is experienced as unpleasant, the sound is judged as dissonant. This can be illustrated with the simple example of the interval. An interval is a combination of two tones of different frequencies. The difference in frequencies between the two tones of an interval, also referred to as interval size, is often expressed as either semitones or multiples of

• **Marcel R. Zentner**, Department of Psychology, University of Geneva, 9 Rte de Drize, CH-1227 Geneva-Carouge; e-mail: zentner@fapse.unige.ch.

INFANT BEHAVIOR & DEVELOPMENT 21 (3), 1998, pp. 483–492
Copyright © 1998 Ablex Publishing Corporation

ISSN 0163-6383
All rights of reproduction in any form reserved.

semitones. The semitone is the smallest difference in Western musical scales and corresponds to a frequency ratio of the two tones of approximately 1.06.²

Although consonance and dissonance have been described as referring to a subjective experience, adults asked to rank intervals in accord with their pleasantness, produce similar results across varied studies. These studies have been reviewed by Schellenberg and Trehub (1994). Adults judge as most consonant either the octave (difference of 12 semitones), the fifth (7 semitones), or the major third (4 semitones). Adults judge as most dissonant the minor second (difference of 1 semitone). These robust facts suggest a lawfulness to the judgement of consonant or dissonant intervals.

This regularity has motivated a search for the *objective* acoustic properties that characterize consonance and dissonance. A popular explanation of consonance and dissonance of

dyadic intervals, put forward by Helmholtz (1954), is that tone pairs are judged consonant when the maximum number of upper harmonics match. Harmonics are the frequency values of individual pure-tone components. These harmonics are usually integer multiples of the fundamental frequency. For example, a complex tone with a fundamental frequency of 100 Hz has of upper harmonics of 100, 200, 300, 400, 500 Hz, and so on. To maximize the number of matching harmonics, the frequency ratios between the fundamentals should be expressible as a ratio of small whole numbers (2:1, 3:2, 4:3, 5:4, 6:5 etc.). Figure 1 presents the example of three intervals with decreasing degrees of consonance as a function of decreasing *simplicity of frequency relations* between the harmonics and the fundamental.

This idea was refined by Plomp and Levelt (1965) who defined a *critical band* for two pure sinusoidal waves as the minimum fre-

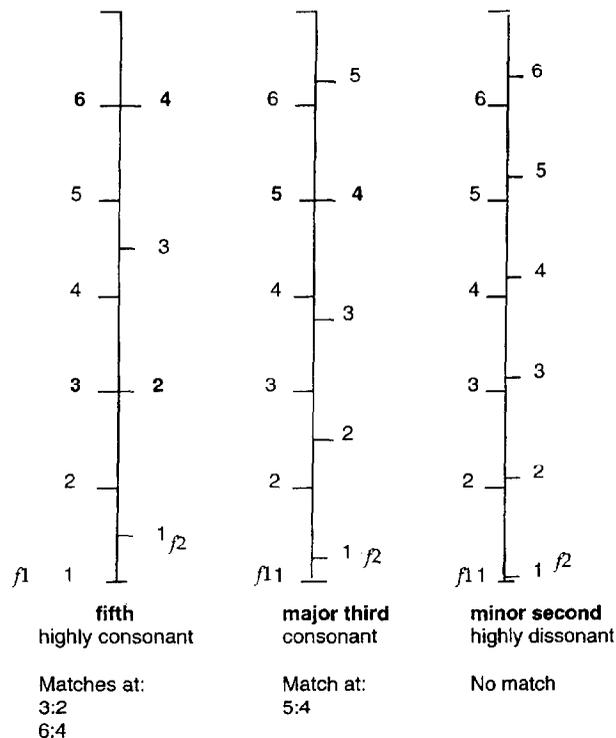


FIGURE 1

The upper harmonics of representative musical intervals and their superposition. Matches of upper harmonics are bold.

quency separation between them that does not give rise to dissonant effects. The critical band is roughly a little lower than a minor third (3 semitones). This means that, for pure sounds, any interval except a full tone (2 semitones) is not dissonant. This generalization does not hold for more complex sounds because they may have pairs of upper harmonics that fall in the critical band. The theory implies that complex tones related by complex frequency ratios are more likely to give rise to dissonant effects not because of the complexity of frequency ratios in itself but because of the greater number of overlapping critical bands among adja-

believed that consonance judgments were the psychological result of physical-acoustical laws operating on *inborn* properties of the auditory system (Plomp & Levelt, 1965; Terhardt, 1984), there is, at present, no proof of this biological preparedness. One could argue that psychoacoustic laws, as posited by Helmholtz and others, could reflect an acquired, rather than an inborn, form of auditory processing. Indeed, some investigators have argued that consonance judgments are acquired competences based on exposure to the music of a particular culture (Francès, 1988; Lundin, 1985). This view is popular among modern music theorists and composers who adopt a sceptical stance toward an absolute notion of consonance (e.g. Boulez, 1971). This "nurture" view can be traced to Schönberg's declaration that the concept of consonance has no useful meaning. Not only did Schönberg proclaim the "emancipation of dissonance," but he also proscribed an elimination of consonances (Schönberg, 1984).

Different strategies can be chosen to probe the origins of our perception of consonance and dissonance. One strategy is to compare judgments of subjects from cultures with different musical systems. If consonance judgments are primarily a function of exposure we should find variability in judgments across cultures. Unfortunately, evidence from such

cross-cultural studies is scant and ambiguous. Consonance judgements were invariant among Americans and Japanese (Butler & Daston, 1968), but dissimilar among Canadians and Indians. Indians showed greater tolerance toward dissonant intervals (Maher, 1976). Hence, the cross-cultural study of consonance preferences has not yet proved to be informative. Although additional evidence would be of interest, it is difficult, if not impossible, to find cultures that are totally isolated from Western music.

A second strategy is to study the reactions of animals to consonance and dissonance.

containing a loudspeaker and two pedals. When the rat pressed one pedal a consonant chord was heard, pressing the other pedal was followed by a dissonant chord (unfortunately, the chords were not specified in the article). The association between the consonant and dissonant chords and the position of the pedals was randomly distributed among animals. Each rat spent 15 min a day in the test chamber in a stipulated sequence for three weeks. A preference for each animal was defined as the difference between the number of presses on the two pedals. The results show that the rats developed a *consonance preference*. After an initial period of small and unreliable differences, the number of presses on the consonance producing pedal increased, and was at the end almost twice the number of presses on the dissonance producing pedal (Borchgrevink, 1975). Although a generalization from rats to humans is speculative—and the prior musical experience of the rats is not known—this study provided support for a possible biological basis for consonance preference.

In a more recent study European Starlings (*Sturnus Vulgaris*) were trained to peck at one key when a consonant chord was presented and at another key when a dissonant chord was presented. The birds generalized the response to a new pair of consonant and dissonant chords (Hulse, Bernard & Braaten, 1995). This

result suggests that consonance and dissonance may be an important cue for auditory communication among songbirds.

The problems with conducting cross-cultural studies and the interpretation of animal work motivate a search for other strategies. The study of young infants provides an opportunity for the exploration of auditory predispositions in humans. Although there is a body of literature on infants' music perception, studies typically focused on infants' processing capabilities rather than preferences and/or affective responses (Trehub & Trainor, 1993). Moreover, these studies of infants' interval perception have examined responses to melodic intervals where tones are presented successively, rather than harmonic intervals where tones are presented simultaneously.

One of the exceptions is a study by Crowder, Reznick and Rosenkrantz (1991) originally designed to test 6 month-old infants' reactions to major and minor chords. Although there were no differences to major and minor modes, the introduction of a dissonant chord seemed to provide reliable differences with prolonged visual fixation on the loudspeaker to the consonant chords. Because the study was not designed to test infants' reactions to consonance and dissonance, the number of infants in the consonance/dissonance condition was very small ($n = 9$).

In a recent experiment by Schellenberg and Trehub (1996) nine month old infants were able to detect changes from harmonic intervals with simple frequency ratios to intervals with complex frequency ratios, but failed to detect changes from complex to simple frequency ratios. The particular intervals used were the fourth and the fifth (= simple frequency ratios) as opposed to the tritone (= complex frequency ratio). The authors concluded that infants possessed a greater facility to discriminate intervals with simple compared to intervals with complex frequency ratios. This result is interesting as it indicates that the distinctiveness and memorability of certain intervals may underlie a preference.

The present experiment was designed specifically to test the hypothesis of an innate preferential bias favoring consonance over dissonance. Unlike studies that used isolated pairs of intervals, infants were presented with actual music. Two different, unfamiliar melodies were composed with a synthesizer, and a consonant and dissonant version was created for each melody. Because our interest was in the infants' preference, and not in perceptual capability, the dependent measures of visual fixation of the music source, motor activity, vocalization, fretting and turning away from the music source were all coded as relevant indicators. It was expected that the infants would be more attentive and show less frequent signs of distress to the consonant compared to the dissonant melodies.

METHOD

Participants

Thirty-two full term, healthy 4 months old infants (16 males, 16 females) participated in the experiment (age range 16.7-21.1 weeks, $M = 17.9$ weeks). Nine additional infants were tested but excluded because they became fussy (7) or because of experimenter error (2). The parents were middle-class caucasian residents of the Boston, Massachusetts metropolitan area and were recruited through the birth records of local city halls.

Stimuli

Each infant was presented with two different melodies, each 35 sec duration, in a consonant and a dissonant version for a total of 4 trials. The consonant and dissonant versions of the two melodies were identical with regard to parameters such as pitch, tempo, rhythm, timbre, contour. The dissonant version was composed in parallel minor seconds (interval size of 1 semitone), the consonant version in parallel major and minor thirds (interval size of 4 and 3 semitones respectively), using only two synthetic

and was covered with an attractive pattern of black and white concentric circles. A video camera recorded the infant's behavior. There was a wall in front and to the left of the infant. A research assistant and the parent, who remained in the room, sat behind the infant, outside of the infant's visual field. Initially, there was a 35 sec quiet baseline followed by the 4 melodies with a quiet interval between each melody of 8 sec. The order of the presentation of the melodies was counterbalanced (Latin Square design). The music was delivered at a SPL of 60dB(A).

Measures

The videotape record was coded for (1) fixation time directed at speaker (2) motor activity (3) turning away from the speaker (4) vocalization and (5) fretting. The fixation time

variable was total fixation time during the 35 sec of each trial, the duration of the initial fixation following the onset of each trial, and the duration of fixation on speaker during the interstimulus intervals. Motor activity was defined as the total duration of (a) flexion of one or both arms and legs of at least 60 degrees, (b) movement of one or both arms and legs by at least 60 degrees to the right or left and more than 2.5 cm in vertical direction. Turning away was defined as the frequency of turning the head away from the pattern of concentric circles by 90 degrees or more. Vocalizations, non-distress vocal sounds, was defined as the frequency of vocalizations per trial. There had to be a quiet period of at least 3 sec between two vocalizations in order to code them as separate. Fretting, defined as distressed vocal sounds, was coded in a similar manner to vocalizations.

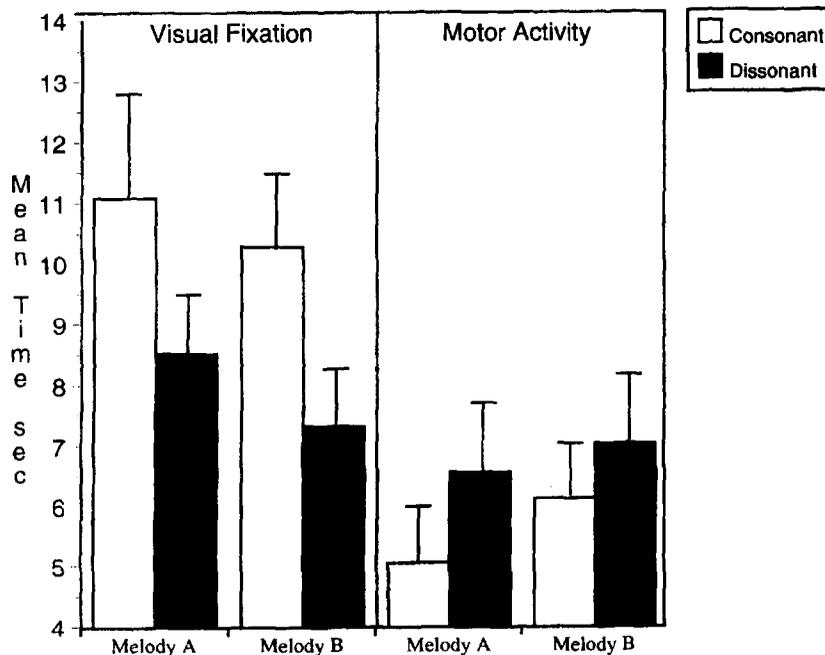


FIGURE 3

Left side: Mean total visual fixation time to music source during the consonant and dissonant versions of two melodies. Right side: Mean time of motor movement during the consonant and dissonant versions of the two melodies. Error bars represent standard errors.

Reliability of the coders was assessed by having two persons independently code these variables for 20% of the subjects. Pearson product moment correlation coefficients were .94 for fixation time, .91 for motor activity, .96 for turning away and .89 for vocalizations and for fretting.

In addition, each parent filled out a questionnaire about listening habits, musical background and education, and their memories of the type of music the infants' might have heard at home.

RESULTS

Visual Fixation and Motor Activity

A repeated measures analysis of variance, with sex and order as the between-subject variables and melody (A or B) and mode (consonant or dissonant) as the repeated measure, was implemented for each dependent variable. A square root transformation was performed because the scores were not normally distributed. The repeated measures ANOVA yielded a significant main effect of musical mode on total fixation time [$F(1,24) = 11.50$ $p < 0.005$], but also on duration of initial fixation [$F(1,24) = 4.72$ $p < 0.05$] and fixation time during the 8 sec intervals [$F(1,24) = 5.30$ $p < 0.04$]. The infants looked significantly longer at the speaker for the consonant, compared with the dissonant, version of each melody (see Figure 3).

Infants also moved significantly more during the dissonant compared with the consonant versions of the melodies [$F(1,24) = 6.64$ $p < 0.02$] (see Figure 3). There was no significant effect for melody (A or B), sex, or order of presentation. There was an unexpected, but significant melody x sex interaction [$F(1,24) = 15.40$ $p < 0.001$]. Girls were more active motorically during melody B, boys were more active during melody A. Melody B had a faster tempo than melody A.

Vocalizations, Fretting and Avoidance

Eight of 32 infants fretted or showed avoidance during the dissonant versions, but neither fretted nor avoided during the consonant versions. No infant fretted or avoided only during the consonant versions but not during the dissonant versions. The probability of this result occurring by chance is $p < 0.01$ by the binomial theorem. Similarly, 7 infants vocalized during the consonant, but not the dissonant versions, while 1 infant vocalized during the dissonant but did not do so during the consonant versions [binomial test: $p < 0.05$].

Musical Background Variables

In order to examine possible effects of the infants' and parents' musical experience we created a new variable by subtracting total fixation time to the dissonant versions from total fixation time to the consonant versions: The larger this difference score, the longer the infant fixated the concentric circles during the consonant melodies. Similarly, we subtracted the total motor score to the consonant melodies from the total motor scores to the dissonant ones: The larger the difference, the more active the infant was during the dissonant conditions. Pearson product-moment correlations were computed between infants' exposure to music at home (average hours/week) and these difference scores for fixation time and motor activity: All correlations were nonsignificant (fixation: $r = -.02$; motor activity: $r = -.10$). Similarly, there were no significant correlations between the infants' fixation time or motor behavior and either the parents' musical background or practice of a musical instrument (fixation: $r = -.22$; $r = -.08$ - motor activity: $r = .19$; $r = -.06$).

DISCUSSION

Infants looked significantly longer at the speaker and were less motorically active when hearing the consonant as compared to the dis-

sonant versions of each melody. Although only a small number of infants vocalized, fretted or turned away, the occurrence of these behaviors was differentially distributed over the consonant and dissonant stimuli. These data support the hypothesis of an innate bias favoring consonance over dissonance.

Two major factors control infants' duration of fixation: preference and discrepancy (Fantz, Fagan, & Miranda, 1975). The dissonant versions of the melodies deviated from what most infants might have heard at home. Therefore they can be treated as discrepant events. If discrepancy was controlling the infants' fixation time we would expect the infants to show increased fixation time during the dissonant versions of the melodies. However, because the infants did not show longer fixations to the dissonant stimuli it is fair to assume that it was a preference, and not discrepancy, that controlled the infants' fixation time.

Differences in motor activity are more difficult to interpret. Increased motor activity reflects arousal, but arousal can be pleasant or distressed. However, because the dissonant versions elicited a combination of increased motor activity *but* shorter fixation times and increased fretting/turning away, it is reasonable to assume that the increased motor activity reflected a distressed rather than a pleasant state. One can also interpret the data by regarding a decrease in motor activity as a measure of interest. In this case the reduced motor activity during the consonant versions leads to the same conclusion as the data on fixation time. Fretting is an obvious manifestation of unpleasantness. In sum, these results reveal a pattern suggesting that the human infant may be born with a biological preparedness that makes consonance perceptually more attractive than dissonance.

Some limitations of our study are acknowledged. The primary limitation is a function of the way consonance and dissonance were operationalized. Although the melodies we created were based on massive evidence from adult listeners who judge thirds (major and minor) to be much more consonant than minor

seconds, the two types of intervals also differ in terms of interval size. Thus, one could argue that the differential behaviors to the two melodies has nothing to do with consonance/dissonance, but are due to differences in size of interval.

However, size of complex-tone intervals (as used in our study) is not associated with preference in adults. Sometimes large intervals are judged as unpleasant (seventh = 10-11 semitones); sometimes they are rated as consonant (sixth = 8-9). The same is true for small (third = 4-5 vs. second 1-2) and medium-sized intervals (fourth = 5 and fifth = 7 vs. tritone = 6). Clearly, then, in adults preference for intervals is not a function of size, but of frequency ratio. Should this be different for infants? In the study by Schellenberg and Trehub cited earlier, the intervals were the fourth and the fifth (= simple frequency ratios), as opposed to the tritone (= complex frequency ratio). When infants were presented with standard interval patterns they were better at discriminating a change from a pattern composed of the fifth or the fourth to the tritone than vice versa. Because the tritone has an interval between the fourth (smaller) and the fifth (larger), frequency ratio, and not interval size, has to be responsible for the results. Further, infants as well as adults, were better able to discriminate changes from a standard consonant harmonic interval to a dissonant harmonic interval than to another consonant interval, even though the change in interval size to the consonant interval was twice as large as the change to the dissonant interval. In this study, dissonance was more important than interval size in mediating interval discrimination (Schellenberg & Trainor, 1996). In sum, although we cannot conclusively rule out the alternative hypothesis, it is less likely than the one proposed.

Another limitation is that seconds and thirds represent only one aspect of consonance and dissonance. Therefore, we must be cautious in generalizing our findings to other forms of consonance and dissonance. More ambiguous forms of consonance and disso-

nance might be subject to a much stronger cultural impact.

Our primary objective in the present investigation has been to examine the hypothesis of an innate bias favoring consonance over dissonance. We can offer no explanations, only speculation, on the nature of the innate bias that renders consonance perceptually more attractive than dissonance. A number of investigators contend that, intervals with matching upper harmonics (= consonances) and intervals with non-matching upper harmonics (= dissonances) create distinct firing patterns in the auditory neural network. This idea, put forward by Boomsalter and Creel (1961) and Moore (1989) (see also Schellenberg & Trehub, 1994), is that complex tones that are related by simple frequency ratios produce similar firing patterns in the auditory network because of the relatively frequent match of upper partials (see Figure 1). The absence of shared neural circuits presumed when there is processing of dissonances might be one reason for experiencing intervals with complex frequency ratios as dissonant.

Another explanation centers on the concept of critical bandwidth. In this view, the auditory neural network has a limited ability to resolve different tones (and its upper harmonics) that are too proximate in pitch. Tones that are very proximate, but not identical, in frequency can not be effectively processed by the basilar membrane: hence the excitation patterns overlap. These fluctuations result in a perception of roughness or dissonance. Because the minor seconds used in the dissonant versions of our melodies are characterized by more complex frequency ratios and more overlapping critical bands than the thirds used for the consonant versions of the melodies, the two hypotheses are equally plausible.

These explanations are reminiscent of Galileo's views for he wrote in 1638 (1963, p. 100):

Agreeable consonances are pairs of tones which strike the ear with a certain regularity; this regularity consists in the fact that the pulses delivered by the two tones, in the

same interval of time, shall be commensurable in number, so as not to keep the ear drum in perpetual torment, bending in two different directions in order to yield the ever discordant impulses.

Acknowledgment: A portion of this work previously appeared in *Nature* (383, p. 29, 1996). Preparation of this paper was supported by a grant from the John D. and Catherine T. MacArthur Foundation Network on Psychopathology and Development and a fellowship from the Swiss National Science Foundation awarded to the first author. We are grateful to Curt Stallman for his help with the creation of the stimuli, to Emily Hoffman, Steve Most and Meredith Rowe for helping us with testing and coding, and to Carol Krumhansl and Sandra Trehub for their helpful comments at different stages of this research.

NOTES

1. Marcel R. Zentner is now at the Department of Psychology at the University of Geneva.
2. The pairs of tones can occur in succession or simultaneously. The former type of interval is called sequential or melodic, the latter is referred to as harmonic. Although consonance and dissonance can refer to a sequence of single tones, they usually refer to harmonic intervals. It is only in this latter sense that we are concerned with consonance and dissonance here.
3. The *exclusive* use of thirds would result in the violation of tonality, introducing another type of dissonance. Thus, in the few instances where it was unavoidable the third was replaced by its natural complement, the sixth, which has a very similar consonance quality. Where the third was replaced by the sixth in the consonant version, in the dissonant version the natural complement of the second, the seventh, was used in order to match the interval size of the consonant version (see Figure 2).

REFERENCES

- Boomsalter, P. & Creel, W. (1961). The long pattern hypothesis in harmony and hearing. *Journal of Music Theory*, 5, 2–31.

- Borchgrevink, H. M. (1975). Musikalske akkord-preferanser hos mennesket belyst ved dyreforsøk. [Musical chord preferences in humans as demonstrated through animal experiments]. *Tidsskrift for den Norske Laegeforening*, 95, 356–358.
- Boulez, P. (1971). *Boulez on music today*. London: Faber and Faber.
- Butler, J. W. & Daston, P. G. (1968). Musical consonance as musical preference: A cross-cultural study. *Journal of General Psychology*, 79, 129–142.
- Crowder, R. G., Reznick, J.S. & Roscnkrantz, S. (1991). Perception of the major/minor distinction. Preferences among infants. *Bulletin of the Psychonomic Society*, 29, 187–188.
- Fantz, R. L., Fagan, J. F., Miranda, S. B. (1975). Early visual selectivity. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition: Vol 1. Basic visual processes* (pp. 249–341). New York: Academic Press.
- Francès, R. (1988). *The Perception of Music* (W. J. Dowling, Trans.). Erlbaum: Hillsdale.
- Galileo Galilei (1963 /1638). *Dialogues concerning two new sciences* (H. Crew & A. de Salvio Trans.). New York: McGraw-Hill.
- Helmholtz, H. v. (1954 /1863). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellism, Trans.) . New York: Dover.
- Hulse, H. S., Bernard, D. J. & Braaten, R. F. (1995). Auditory discrimination of chord-based spectral structures by European Starlings (*Sturnus vulgaris*). *Journal of Experimental Psychology: General*, 124, 409–423.
- Lundin, R. W. (1985). *An objective psychology of music*. (3rd ed.). Malabar, FL: Krieger.
- Maher, T. F. (1976). "Need for resolution" ratings for harmonic musical intervals: A comparison between Indians and Canadians. *Journal of Cross Cultural Psychology*, 7, 259–276.
- Moore, B. C. J. (1989). *An introduction into the psychology of hearing* (3rd ed.). London: Academic Press.
- Plomp, R. & Levelt, W. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38, 548–560.
- Schellenberg, E. G. & Trainor, L. T. (1996, June). *Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: Tests of adult and infant listeners*. Paper presented at the 10th Biennial International Conference on Infant Studies, Providence, Rhode Island.
- Schellenberg, E. G. & Trehub, S. E. (1994). Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review*, 1, 191–201.
- Schellenberg, E. G. & Trehub, S. E. (1996). Natural musical intervals. Evidence from infant listeners. *Psychological Science*, 7, 272–277.
- Schönberg, A. (1984). *Style and idea*. Berkeley: University of California Press.
- Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception*, 1, 276–295.
- Trehub, S. E. & Trainor, L. J. (1993). Listening strategies in infancy: The roots of language and musical development. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 278–327). London: Oxford University Press.

12 December 1996; Revised 28 October 1997 ■