

Change Detection in Multi-Voice Music: The Role of Musical Structure, Musical Training, and Task Demands

Edward J. Crawley
Marywood University

Barbara E. Acker-Mills
Auburn University at Montgomery

Richard E. Pastore
State University of New York at Binghamton

Shawn Weil
Ohio State University

This study evaluated the relationship between primitive and scheme-driven grouping (A. S. Bregman, 1990) by comparing the ability of different listeners to detect single note changes in 3-voice musical compositions. Primitive grouping was manipulated by the use of 2 distinctly different compositional styles (homophony and polyphony). The effects of scheme-driven processes were tested by comparing performance of 2 groups of listeners (musicians and nonmusicians) and by varying task demands (integrative and selective listening). Following previous studies, which had tested only musically trained participants, several variables were manipulated within each compositional style. The results indicated that, although musicians demonstrated a higher sensitivity to changes than did nonmusicians, the 2 groups exhibited similar patterns of sensitivity under a variety of conditions.

The processes by which we achieve a meaningful parsing of the auditory scene remain the topic of much speculation, debate, and research. Bregman (1990) proposed that a meaningful parsing of the auditory scene depends on two conceptually distinct processes: bottom-up (stimulus-driven or primitive) and top-down (scheme-driven or secondary). Most of the existing research on auditory scene analysis (ASA) has focused on the primitive processes by actively seeking to limit the influence of scheme-driven processes. One consequence of this strategy is the use of relatively simple stimuli that would seldom be encountered outside the laboratory (e.g., Barsz, 1988; Bregman & Campbell, 1971; Lackner & Goldstien, 1974). This research strategy has been very effective in identifying many of the important parameters for stimulus-driven grouping, and this knowledge provides the background necessary for studying the seldom investigated scheme-driven processes.

The present study evaluated the relationship between primitive and scheme-driven grouping by comparing the ability of listeners

to detect single note changes (the substitution of a different note) in musical compositions. Primitive grouping was manipulated through use of contrasting compositional styles (homophonic and polyphonic). The effects of scheme-driven processes were tested by comparing the performance of two groups of listeners (musicians and nonmusicians) and by varying task demands (integrative and selective listening).

Primitive Processes and Compositional Style

The primitive processes are assumed to reflect relatively fixed properties of the auditory system and thus should not differ dramatically as a function of an individual's experience with the auditory environment. The major stimulus dimensions used by the primitive processes to analyze the auditory scene are temporal proximity and frequency similarity (e.g., Beauvois, 1998). Specifically, components of a complex auditory signal that are similar in their temporal or frequency characteristics tend to group together, forming a cohesive (integrated) auditory object or stream that is perceptually segregated or removed from the remaining auditory signal (the residual).

The results of the primitive processes impact a listener's ability to extract information from an auditory signal. For example, synchronous tones tend to integrate or perceptually fuse, producing a temporally (vertically) coherent signal, especially if the tones are in a good harmonic relationship (Bregman, 1990, pp. 320, 681–683; Dewitt & Crowder, 1987). In this case, global properties or emergent features become more salient than the individual tones. Thus, homophonic musical pieces, which are generally defined by synchrony of notes across frequency regions (voice), should support a vertical organization and therefore bias a listener toward perceiving a series of chords (wholistic or integrative listening).

Stimuli that contain temporal asynchrony of notes across frequency regions, as is typical in polyphonic music, will be more

Edward J. Crawley, Department of Psychology, Marywood University; Barbara E. Acker-Mills, Department of Psychology, Auburn University at Montgomery; Richard E. Pastore, Department of Psychology, State University of New York at Binghamton (SUNY-Binghamton); Shawn Weil, Department of Psychology, Ohio State University.

Portions of this research were presented at the 133rd Meeting of the Acoustical Society of America, State College, Pennsylvania, June 1997, and at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, November 1997. This research was conducted in the Auditory Cognition Laboratory at SUNY-Binghamton and was supported in part by Air Force Office of Scientific Research Grant F49620-93-10327 to Richard E. Pastore. We thank John Rothgab of SUNY-Binghamton's Music Department for providing the original polyphonic stimuli used in the experiment.

Correspondence concerning this article should be addressed to Edward J. Crawley, Department of Psychology, Marywood University, 2300 Adams Avenue, Scranton, Pennsylvania. E-mail: crawley@ac.marywood.edu

likely to produce horizontal (within-voice) integration. This horizontal integration should produce perceptually isolated frequency regions, with weakened information about vertical (between-voice) relationships. Thus, polyphony should create a situation in which listeners can monitor individual voices (i.e., focused or analytic listening), but the ability to listen wholistically (i.e., simultaneously across frequency regions) should be greatly compromised (Sloboda, 1985).

Primitive Processes and the Present Stimuli

The six basic stimuli in the present study are three-voice polyphonic (fourth species counterpoint) and homophonic compositions designed by Acker and Pastore (1996b) to represent distinct examples of homophony and polyphony. Figure 1 presents one homophonic piece (with the melody in the high voice), with the remaining stimuli presented in the Appendix. For purposes of the present study, a musical voice was defined as a sequence of notes occurring within a relative and limited frequency region or location (i.e., high, middle, or low voice location). The three voices are easily seen by their representation on different staves in the Appendix. Although the frequency ranges of neighboring voices were not exclusive, the relative frequency height within any particular measure continued to define voice location. Another aspect of any multivoice piece, including the present stimuli, is the presence of information in both vertical (harmonic) and horizontal (melodic) dimensions. The vertical and horizontal dimensions are apparent from the spatial arrangement of the notation. The horizontal dimension refers to the within-voice relationships that may produce a melody as notes are presented over time. The vertical dimension refers to the harmonic relationships between voices at any given time; in this case the simultaneous notes produce major chords.

Figure 1 and the first three examples in the Appendix present the homophonic pieces used in the present studies. These homophonic pieces were defined by complete synchrony across voices and thus were expected to promote a strong vertical grouping. The polyphonic stimuli used in the present studies (see the bottom half of the Appendix) were expected to promote a horizontal grouping through the use of asynchronous onset of notes in one of the nonmelody voices relative to the melody.

Although the stimuli in the current experiments were created to be distinct examples of polyphony and homophony and were expected to promote either vertical (homophonic compositions) or horizontal integration of voices (polyphonic compositions), neither type of grouping is ever absolute in musical compositions. A

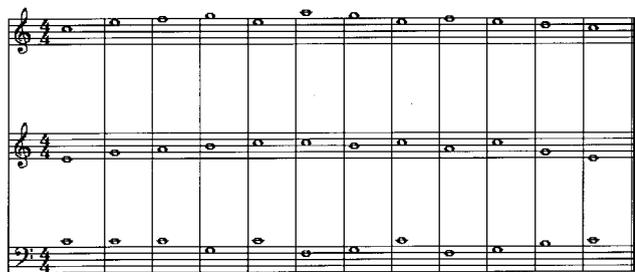


Figure 1. The homophonic composition containing the melody in the high voice, one of the six standard stimuli used in the present experiments.

melody (resulting from horizontal grouping) can be heard even in a strongly homophonic composition. In polyphonic compositions, the secondary voices are typically written to create some harmonic coherence with the primary voice. Within a compositional style there can also be considerable variation in grouping resulting from the composition's specific arrangement. For example, the results of several studies (Acker & Pastore, 1996a, 1996b; Palmer & Holleran, 1994) have shown that within a specific compositional style a number of different factors, such as melody location, also influence the ability of a listener to detect changes in a multivoice composition.

Scheme-Driven Grouping

Scheme-driven processes, by definition, develop through learning and experience with the auditory environment. Thus, because they reflect learning and stored representations, the types and availability of scheme-driven processes should differ as a function of an individual's experience or level of expertise. It has also been argued that schemes, following the operation of the primitive processes, can alter the organization of the acoustic signal through expectations and voluntary attention (Bregman, 1990). The present study sought to determine the relative influence of schemes by comparing the ability of listeners with differing levels of musical experience (musicians vs. nonmusicians) to choose a mode of listening (selective vs. integrative) that was most efficient for the given task. To the extent that schemes can alter the perceptual organization resulting from the primitive processes, task performance should show independence from the primary grouping on the basis of the physical characteristics of the auditory stimuli. Previous studies addressing the effects of musical training on the ability to choose to listen in either an integrative or selective manner have provided conflicting results.

Several studies have found that musical training may allow listeners to overcome the effects of stimulus-driven grouping in an integrative listening task (Acker & Pastore, 1996a; Palmer & Holleran, 1994). Conversely, other studies have indicated that the ability to choose to listen either selectively or integratively ("attentional flexibility") may be constrained by primitive processes (Acker & Pastore, 1996b; Botte, Drake, Brochard, & McAdams, 1997; Jones, Jagacinski, Yee, Floyd, & Klapp, 1995). The apparent discrepancy in results is difficult to resolve because these previous studies have used differing stimuli and either have not compared the abilities of musicians and nonmusicians (e.g., Acker & Pastore, 1996a, 1996b; Palmer & Holleran, 1994) or have not compared these two groups in tasks that require the detection of changes in musical structure (Jones et al., 1995).

In a set of experiments that were conceptually similar to those in the present study, Jones et al. (1995) tested the attentional flexibility of musicians and nonmusicians in a task that required the detection of a temporal irregularity. Jones et al. found that musicians, although better able to detect rhythmic changes, did not demonstrate an enhanced flexibility for perceptual organization. That is, the pattern of performance for musicians, which was similar to that of nonmusicians, was largely controlled by stimulus-driven effects rather than intentions to listen in either an integrative or focused manner. There are several possible interpretations of these results.

First, formal training may simply build on, and allow more efficient use of, processes that are available to both musicians and nonmusicians. This more efficient processing for musicians would increase the signal-to-noise ratio and result in a greater consistency of responding. Alternatively, because the stimuli used by Jones et al. (1995) lacked the types of harmonic and melodic structure that characterize most musical compositions, it is possible that the stimuli did not provide an opportunity for musicians to apply their schemes or did not require musicians to rely heavily on the use of schemes. Thus, it is unclear whether the lack of qualitative difference in performance for musicians and nonmusicians resulted from the presence of similar processes in both groups of listeners or resulted simply because previous tasks and stimuli did not require musicians to rely on their schemes.

To evaluate these possibilities, we used relatively complex but musically valid stimuli in the present experiments, with the expectation that the musical validity would allow for the use of schemes and that the relative complexity would require a reliance on these schemes. The idea was to provide a medium in which musical schemes should be able to operate, and then create a situation in which the task requirements were either relatively compatible or incompatible with the grouping that results from the operation of the primitive processes. When the task requirements are relatively incompatible with the output of the primitive processes, more effort should be required to perform the task, and thus the effects of scheme-driven processes should become apparent.

Change Type: Chord-Related and Chord-Unrelated Changes

The present experiment incorporated a manipulation that has been demonstrated to be an important influence on musicians' detection performance. Acker and Pastore (1996a, 1996b) demonstrated that musicians find it more difficult to detect diatonic note changes that are *chord-related* relative to those that are *chord-unrelated* when the task requires the detection of any single note change in the compositions. Chord-related changes are created by changing one note of a major chord (e.g., C major consists of the notes C, E, and G) to another chord tone (e.g., the G changes to a C). Chord-unrelated changes are created by changing one chord tone to a diatonic note that is not the root, third, or fifth of the chord (e.g., the G changes to a B). Thus, for chord-unrelated changes, note changes remain within the key of the piece but fall outside of the harmonic structure of the original chord. Additionally, because of the constraints on the nature of the introduced changes, chord-related changes were, on average, associated with a larger frequency change than chord-unrelated changes. Thus, the finding of differential sensitivity to these two classes of note changes, with chord-related changes detected less often than chord-unrelated changes, indicates that, at least for trained musicians in an integrative listening task, musical relatedness is relatively more important than is the absolute change in frequency.

The Present Study

In both experiments of the present study, listeners were asked to detect single note changes that were introduced into polyphonic and homophonic musical compositions. In Experiment 1, listeners were familiarized with a three-voice musical composition and then

were required to detect the presence of any change within subsequent presentations of the musical piece. Detection of any change should be easier if the listener can simultaneously monitor or integrate information contained across voices (e.g., by monitoring changes in simultaneous intervals). Thus, in the homophonic context, because the task demands and stimulus-driven grouping were compatible, detection of any change was expected to be easier than in the polyphonic context. Specifically, the primitive processes should tend to produce integration and the participants should be "trying" (a secondary process) to hear the stimuli in an integrative manner. A different pattern of results would be expected if the physical properties of the stimuli were changed such that they were incompatible with an integrative listening task.

The asynchrony between voices in the polyphonic composition should lead to greater within-voice integration, making it easier to monitor an individual voice but making it more difficult to simultaneously monitor all three voices. In this case, the output of the primitive processes is relatively incompatible with the task demands. Thus, if musical training allows more perceptual flexibility, musicians, relative to nonmusicians, should be better able to overcome the relative incompatibility between the output of the primitive processes and the task requirements.

In Experiment 2, listeners were first familiarized with the melody independent of the accompaniment. Following this familiarization phase, the listeners were presented with the three-voice composition (melody plus accompaniment) and were asked to detect changes occurring in the melody and to ignore changes occurring in the accompanying voices. Thus, the task in Experiment 2 can be characterized as a focused attention task in which listeners are required to monitor (focus on) only a single voice and to ignore the other voices. Given these task requirements, the effects of compositional style should be the opposite of those observed in Experiment 1. That is, the greater degree of between-voice integration in the homophonic context should make it more difficult for listeners to attend to a single voice (or filter out information from the other voices).

Weakening the between-voice integration, as in the polyphonic context, should increase a listener's ability to process the voices independently. In Experiment 2, within the polyphonic context the primitive and scheme-driven processes were compatible, and thus we expected melody-change detection to be higher relative to that in the homophonic context. If the effects of musical training increase the ability to focus attention and to switch attention at will (Sloboda & Edworthy, 1981), musicians should not only perform better than nonmusicians but again should also be less affected by changes in compositional style.

Thus, although the musical compositions were identical in the present experiments, the task demands differed. If musical experience leads to increased attentional flexibility (i.e., the ability to reorganize or compensate for the output of the primitive processes), then the performance of the musicians relative to the nonmusicians should be less dependent on the underlying compositional structure in both experiments. On the other hand, if musical training simply builds on processes shared by both musicians and nonmusicians, then although musicians may consistently demonstrate a higher overall level of performance, the pattern of performance for the two groups of listeners should not be significantly different. In either case, whether or not different patterns of

performance would be observed, the present studies would provide insights into the nature of the effects of formal music training.

Experiment 1: Integrative Listening

Experiment 1 represents an important extension of several previous studies that have used the same paradigm but have tested only trained musicians (Acker & Pastore, 1996a; Palmer & Holleran, 1994). The results of these previous studies are consistent with the view that change detection for musicians is not highly dependent on the underlying compositional style.

Although the results of the Acker and Pastore (1996a) and Palmer and Holleran (1994) studies could reflect the operation of musical schemes, it is also possible that the null effect of compositional style reflects a weak manipulation at the level of primitive processing (Acker & Pastore, 1996b). Specifically, both the homophonic and polyphonic compositions used by Palmer and Holleran (and later by Acker & Pastore, 1996a) contained some asynchrony between voices, with the two compositional styles mainly distinguished by the presence of passing tones in the polyphonic context that produced a secondary melody. Although the addition of tones within one frequency region should have increased horizontal integration in that frequency region, thus promoting a different grouping relative to the homophonic composition, other factors probably worked against finding an effect of compositional style. For example, changes in both the polyphonic and homophonic contexts always occurred when all three voices were sounded simultaneously. Thus, the perceptual distinction between compositional styles at the change location was minimal.

The present study was designed to provide a strong manipulation of both primitive grouping and schemas. First, the stimuli for the study were created to represent highly distinct examples of the general compositional styles and, thus, to enhance the effects of primitive processes. Second, as discussed earlier, the study included two groups of listeners (musicians and nonmusicians) who were expected to differ in the availability of musical schemes. Thus, the present experiments tested the generality of previous results by using a set of compositionally distinct stimuli and provided a comparison across level of training. If the null effect of compositional style observed in previous studies (Acker & Pastore, 1996a; Palmer & Holleran, 1994) resulted from a weak manipulation of primitive grouping variables, then, in the present experiment, similar effects of compositional style would be expected for both musicians and nonmusicians. That is, the results would be expected to be consistent with previous studies that have used decidedly nonmusical stimuli (e.g., Botte et al., 1997; Jones et al., 1995). On the other hand, if the null effect of compositional style reflects the operation of musical schemes, then level of training would be expected to interact with compositional style, with the performance of nonmusicians being more highly dependent on compositional style than the performance of musicians.

Method

Participants

A total of 24 participants (12 musicians and 12 nonmusicians) were recruited from the State University of New York at Binghamton (SUNY-Binghamton) population. Musicians were required to have a minimum of 5 years of formal training ($M = 8.6$ years, $SD = 1.2$ years). Although the

background of these musicians was variable, all reported some formal ear training, knowledge of music theory, and participation in high school or university orchestras. All reported that they were currently involved in musical activities. Nonmusicians were selected on the explicit basis of an absence of formal music training or instruction. None of the nonmusicians reported any formal or informal musical training, and therefore the musician and nonmusician participants represented distinctly separate groups. Each participant either received credit toward a research requirement in an introductory psychology class or was paid.

Stimuli

In the present experiments the main distinction between the polyphonic and homophonic stimuli was the degree of synchrony between the three voices. The homophonic pieces contained only whole notes that were synchronous across all three voices. The polyphonic pieces contained synchronous whole notes in two of the voices and tied half notes in the third voice; the onsets of the tied half notes occurred on the third beat of each measure. Thus, the third voice in the polyphonic composition was asynchronous with the other two voices. The melody was never presented in the asynchronous voice.

All the polyphonic stimuli used in the present experiment were based on three examples of counterpoint provided by a SUNY-Binghamton music professor. The examples shared the same overall frequency range and contained the same 12-note melody but differed in the frequency region (voice) of the melody. Thus, within each piece the main melody was contained within a single voice, but across pieces the melody was presented in each voice. To create stimuli for the homophonic condition, Barbara E. Acker-Mills wrote three distinct three-voice harmonizations for the melody used in the polyphonic pieces—one harmonization for the melody in each frequency region. Because the melody, with the exception of voice location, was the same across harmonization, the three homophonic stimuli differed only in terms of chord inversions (i.e., the harmonic progressions were the same). As in the polyphonic example, all three pieces covered approximately the same frequency range. Thus, the six pieces (three polyphonic and three homophonic) each contained the same melody, and each of the three pieces within a compositional style differed in the frequency region (voice) containing the melody.

Single note changes were introduced into these original six pieces at one of three temporal locations (Note 6, 8, or 10). The use of three temporal locations ensured that participants would be more likely to attend to the whole piece rather than to a specific temporal location. These single note changes were characterized as either *chord-related* or *chord-unrelated* changes. A chord-related change involved changing a chord tone (the root, third, or fifth) to another chord tone, thus maintaining the chord identity at the change location. For a chord-unrelated change, the original note was replaced with a diatonic (within the key of the piece) note that was not the root, third, or fifth of the chord at that location. Changes that would produce an interval of a second or seventh were not permitted because these intervals are particularly dissonant. Furthermore, no changes were greater than a fourth (up or down), and no open fifths were created. A total of 18 variations were created for each of the original six pieces (3 temporal locations \times 3 voices \times 2 change types). All stimuli, both the original pieces and those containing a single note change, were created with Encore (1995) sequencing software using a flute timbre. All stimuli were stored as MIDI files for later use by a computer program.

Procedure

Musical experience and musical context (polyphonic vs. homophonic) were manipulated between participants, resulting in four groups (2 contexts \times 2 levels of musical experience). Data for each participant were collected in three experimental sessions across which melody location, and thus the specific musical piece, was varied. Order of presentation of specific musical pieces was counterbalanced across participants.

Before testing began in the first session, participants completed a brief questionnaire about the type and extent of their musical training and the instrument(s) played. The purpose of the questionnaire was to both confirm and quantify the selection of participants meeting the criterion for musicians or nonmusicians. Following completion of the questionnaire, each participant was seated alone in a sound-attenuated chamber. Participants were informed that, for purposes of familiarization, they would hear four repetitions of a short musical piece and that the subsequent task would require them to detect any change (a single note change) in this piece. During the test phase, participants responded using an 8-point scale in which 1 represented *very sure same* and 8 represented *very sure different*. A rating of 4 or less indicated that the participant believed that no change had occurred on the trial (with lower numbers indicating a higher confidence that there was no change). A rating of 5 or more indicated that the participant believed that a change had been present (with higher numbers indicating a higher confidence that a change had occurred). On test trials participants were given 3 s following the offset of the piece to respond before the next trial began. For half of the trials a single diatonic change was present; for the other half the test stimulus was identical to that presented during familiarization. After every seven trials, a single repetition of the original piece was presented as a reminder. During the reminder trials a computer monitor displayed “*Original piece do not respond!*” to ensure that participants knew the original piece was being presented.

Within an experimental session, participants were randomly presented with 54 change and 54 no-change trials in an order that was randomized for each participant and session. On change trials, the type of change (chord related or chord unrelated), the voice containing the change (high, middle, or low), and the temporal location of the change (Note 6, 8, or 10) was randomly selected without replacement from the set consisting of 3 repetitions of each of the 18 variations of the piece. Participants received a short break approximately every 20 min, with the experimental session lasting 1 hr.

Results and Discussion

Table 1 presents a summary of the pattern of significant results.

Confidence Ratings

No-change trials. An analysis of variance (ANOVA) with training, compositional style, and melody location entered as factors revealed no differences in confidence for no-change trials ($M = 3.2$). Despite this low average confidence for *same* judgments (responses were in the middle or “probably same” part of the scale), the average correct-rejection rate (responding “same” when given same) was fairly high (.78). No significant differences or interactions were revealed by an ANOVA conducted with the correct-rejection rates.

Change trials. The present confidence ratings results replicated many of the observations made by Palmer and Holleran (1994), including a null effect of compositional style. As can be seen in Table 1, the pattern of confidence rating was also highly dependent on the participant’s level of musical training. That is, the level of musical training not only resulted in an overall difference in confidence ratings but also influenced patterns of responding across a number of other variables. As shown in Table 1, the pattern of significant results differed dramatically for the confidence ratings and d' .

d' Results

False-alarm rate and hit rate for each condition were calculated for each participant. A false alarm was defined as responding that

Table 1
Summary of Significant Results ($p < .05$) in Experiments 1 and 2

Condition	Experiment 1: Detect any change				Experiment 2: Detect change in melody	
	CR		d'		d'	
	<i>F</i>	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>	<i>df</i>
Training (T)	19.92	1, 20	18.99	1, 20	6.79	1, 10
Compositional style (CS)	—	—	4.88	1, 20	—	—
Change type (CT)	107.11	1, 20	54.97	1, 20	—	—
Change location (CL)	5.51	1, 20	—	—	—	—
Melody location (ML)	5.00	1, 20	—	—	7.92	2, 20
T × CS	—	—	—	—	—	—
T × CL	3.43	2, 40	—	—	—	—
CT × CL	11.98	2, 40	5.48	2, 40	—	—
CS × CT	7.80	1, 20	15.19	1, 20	—	—
CS × ML	—	—	—	—	5.06	2, 20
CT × ML	4.66	2, 40	—	—	—	—
T × ML × CL	2.70	4, 80	—	—	—	—
T × CT × CL	3.65	2, 40	—	—	—	—
T × CS × ML	4.99	2, 40	—	—	—	—
C × CT × CL	—	—	—	—	—	—
C × CT × ML	15.35	2, 40	7.53	2, 40	8.63	2, 20
C × CL × ML	7.18	4, 80	5.36	4, 80	—	—
CT × CL × ML	7.26	4, 80	—	—	—	—
C × CT × CL × ML	—	—	—	—	—	—

Note. Dashes indicate nonsignificant results. Empty cells indicate conditions that were not tested. CR = confidence rating.

a change had occurred when in fact no change had been present. A hit was defined as responding accurately that a change had occurred. The proportion of false alarms and the proportion of hits were converted into Z scores and used to calculate d' , the signal-detection theory measure of sensitivity (Macmillian & Creelman, 1991; Pastore & Scheirer, 1974).

Contrary to what was observed using the confidence ratings measure, the main effect of compositional style was significant when d' was considered. Participants demonstrated higher sensitivity to changes in the homophonic context ($d' = 2.8$) relative to the polyphonic context ($d' = 2.0$). The significant effect of compositional style contrasts with the present confidence-rating results as well as with the findings of Palmer and Holleran (1994), who used a confidence-ratings measure, and Acker and Pastore (1996a), who used d' . The most likely explanation for this later discrepancy is that, relative to the stimuli used in both of the previous studies, the stimuli in the present study were more typical examples of polyphony and homophony (Acker & Pastore, 1996b). The significant effect of compositional style is consistent with theorizing about perceptual grouping in general (e.g., Bregman, 1990) as well as with the findings of several studies using different types of stimuli (e.g., Jones et al., 1995).

In addition to the main effect of compositional style, a number of other significant effects were observed with the d' measure. The significant effect of change type indicates that the harmonic relationship between the single note change and the surrounding musical context is a crucial factor for determining change detectability. Participants were significantly more sensitive to chord-unrelated changes ($d' = 2.73$) than to chord-related changes ($d' = 2.07$). This main effect of change type is consistent with the finding that listeners are sensitive to the "goodness" of simultaneously sounded intervals (e.g., Krumhansl & Shepard, 1979). The effect of change type was also found to interact with the effects of change location and compositional style, and with the Compositional Style \times Melody Location interaction.

The significant Change Type \times Change Location interaction, also observed in both of the previous studies, can be attributed to a slightly larger effect of change type in the middle voice relative to the other two voices. At least two possibilities exist to account for this finding. One is that despite the requirement to integrate information across voices, participants tended to focus their atten-

tion more on the high and low voices. The participants thus would have been more sensitive to subtle changes occurring in the outermost voices. The second possibility is that consonant (related) changes are more likely to become integrated when they are flanked in both higher and lower frequency regions. These two interpretations, although operating differentially at the primitive and secondary levels, are not mutually exclusive.

The pattern of the Change Type \times Compositional Style interaction indicates that, in general, the difference between chord-related and chord-unrelated changes is smaller in the polyphonic context than in the homophonic context. This interaction is consistent with the interpretation that the asynchrony in the polyphonic compositions makes it more difficult for participants to attend to all the voices simultaneously. Difficulty integrating voices could result in listeners' tending to focus on individual voices. This conjecture is supported by the finding of significant Change Type \times Compositional Style \times Melody Location and Compositional Style \times Change Location \times Melody Location interactions.

Figure 2 presents the Compositional Style \times Change Location \times Melody Location interaction. The main feature of this interaction is the low sensitivity to changes occurring in the polyphonic composition with the melody in the high voice. A closer examination of the polyphonic compositions, presented in the Appendix, provides some insight into a possible cause of this finding. Specifically, the polyphonic arrangement with the melody in the high voice is the only one with the asynchronous voice in the middle frequency region. This arrangement could have reduced the vertical coherence of the synchronous voices, thus further decreasing the ability to attend to all the voices simultaneously. The conjecture that the asynchrony in the polyphonic compositions allows for greater segregation (or conversely, that homophony produces greater integration) of voices was further tested in Experiment 2, in which participants were required to detect changes occurring in the melody while ignoring changes occurring in the other voices.

Musical Training and Integrative Listening

Figure 3 presents the Compositional Style \times Change Type \times Melody Location interaction as a function of the level of musical

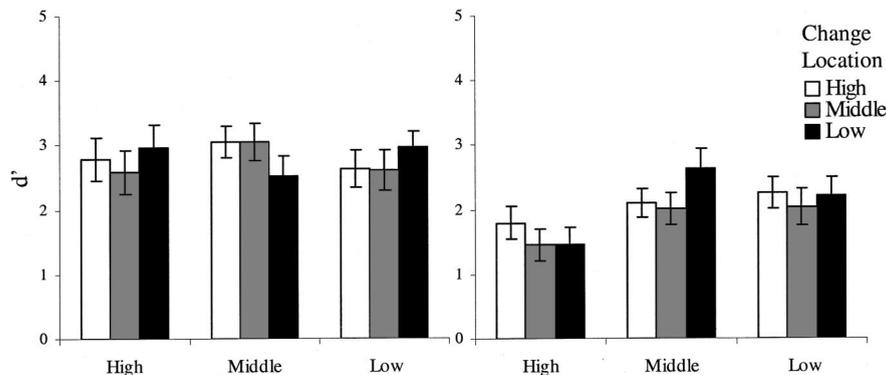


Figure 2. Sensitivity to the presence of a single note change in Experiment 1 as a function of compositional style (polyphonic and homophonic), change location (high, middle, and low voice), and melody location (high, middle, and low voice).

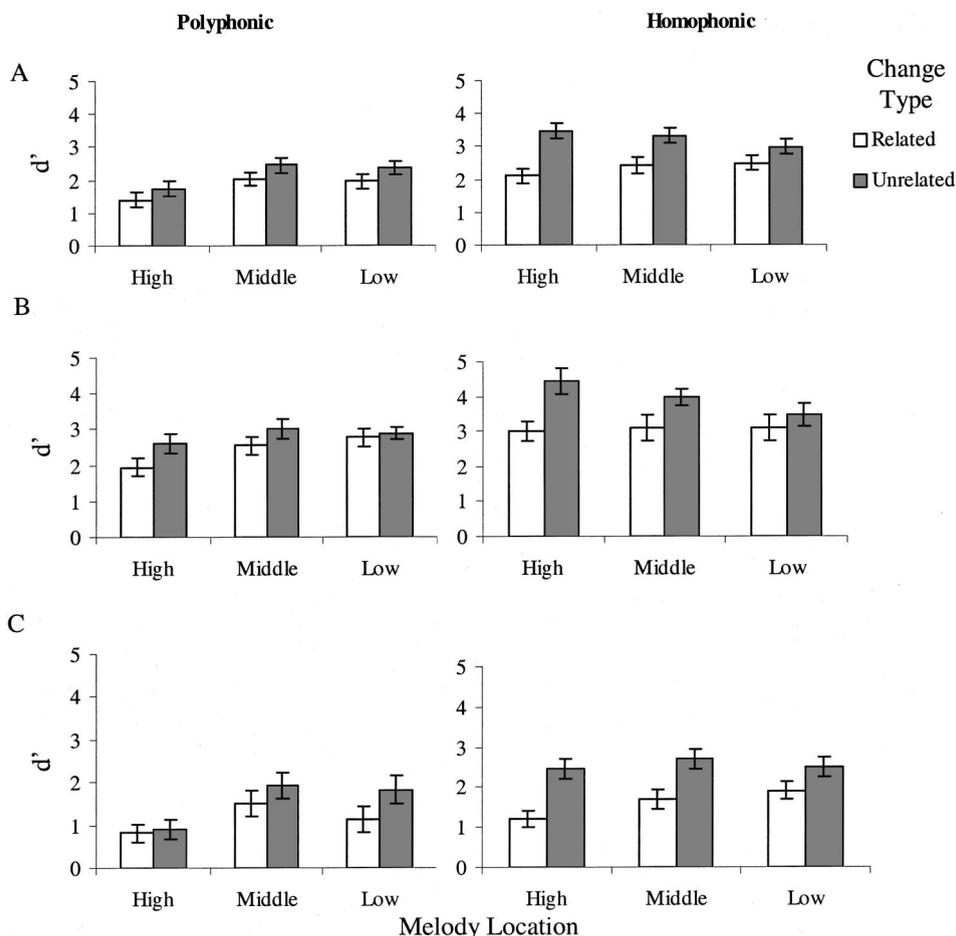


Figure 3. Sensitivity to the presence of a single note change in Experiment 1 as a function of compositional style (polyphonic and homophonic), change type (related and unrelated), and melody location (high, middle, and low voice) averaged across all listeners (Panel A) and separately for musicians (Panel B) and nonmusicians (Panel C).

training. A comparison of the musicians (Panel B) and nonmusicians (Panel C) indicates that although musical training significantly enhanced a listener's performance in the present change-detection task, training did not result in a significant difference in the pattern of performance in this integrative listening task.

Although the musicians in this experiment could be considered only moderately trained (none were professional musicians and average training = 8 years), the current pattern of results is quite consistent with that of Acker and Pastore (1996a), who used only highly trained musician (minimum of 10 years experience). It thus seems likely that increasing the level of musical training, while possibly increasing the overall level of performance, would not lead to any different conclusions.

Experiment 2: Selective Listening

In Experiment 2 listeners were asked to selectively attend to the melody in the three-voice compositions used in Experiment 1, responding only to changes occurring in the melody. Because changes were presented on each trial on the present experiment, the participants had to discriminate changes occurring in the mel-

ody from those occurring in the accompaniment. This defines a focused attention or analytic listening task. To the extent that the stimulus composition or top-down processes allow for segregation of the different frequency regions, discrimination between melody and accompaniment changes should be relatively easy. This task analysis is consistent with Sloboda's (1985) proposal that "focal attention" allows for the detection of subtle changes within the focused stream whereas information pickup from the nonfocused streams is weakened. Some support for this conceptualization is provided by a number of studies of selective attention based on frequency (i.e., pitch) for both simple (e.g., Dai, Scharf, & Buus, 1991) and complex (e.g., Botte et al., 1997; Jones et al., 1995) auditory stimuli.

Compositional Structure and Task Demands

We expected the results of the selective listening task (Experiment 2) to differ in several ways from those observed in an integrative listening task (the present Experiment 1; Acker & Pastore, 1996a; Palmer & Holleran, 1994). If listeners are able to perceptually segregate the target voice (i.e., the melody), change

detection should be similar to that observed in a condition in which the target voice is presented in isolation. Specifically, the polyphonic context should allow greater segregation of voices and thus higher performance relative to the homophonic context (the opposite of what we observed in Experiment 1). Furthermore, as in Experiment 1, the specific arrangement within a compositional style (i.e., melody location) is also likely to impact performance. For example, if our interpretation of the results of Experiment 1 is correct, the polyphonic composition with the melody in the high voice should be most compatible with performing a selective listening task. Thus, in direct contrast to the results of Experiment 1, this specific piece was expected to allow the highest detection rates in the present task.

Level of Musical Training

If formal music training allows for more efficient use of scheme-driven analytic processing, then the current focused attention task should produce differences between musicians and non-musicians as a function of compositional style. This analysis is consistent with the conjectures of Sloboda and Edworthy (1981), who argued that differences between musicians and nonmusicians would most likely be observed in an analytic listening task when using complex stimuli. Thus, the effects of primitive grouping should be stronger for nonmusicians, and consequently, an interaction between training, compositional style, and change type should be observed. Conversely, Jones et al. (1995) found that the compositional structure, not the level of training, was the main determinant of the ability to detect timing variations. If the Jones et al. results generalize to the present stimuli and task, then although musicians should have higher detection rates, the pattern of performance for the two groups should not differ, and thus the effects of training should be consistent with those in Experiment 1.

Method

Participants

Twelve participants (6 musicians and 6 nonmusicians) were recruited from the SUNY-Binghamton population. The 6 musicians reported an average of 10.5 ($SD = 1.0$) years of experience. All the musicians reported some formal training in music theory and ear training. Nonmusicians were chosen explicitly on the basis of having no formal or informal musical training. Participants either received credit toward a research requirement for an introductory psychology class or were paid.

Stimuli

In addition to the set of stimuli used in Experiment 1, the melody in each of the three frequency regions was isolated from the accompaniment and stored as a MIDI file for later presentation. Single note changes, identical to those introduced into the melody in the three-voice context, were introduced into the isolated melody; the distinction between chord-related and chord-unrelated changes does not apply to these single-line "melody-only" stimuli. Thus, six variations (3 temporal locations \times 2 changes) were created for the melody in each frequency region.

Procedure

The procedure used in Experiment 2, identical to that used by Acker and Pastore (1996b), differed from Experiment 1 in several ways. All participants were tested in two experimental sessions, performing the task with

one compositional style in each session and with the order counterbalanced across participants. Each session consisted of four phases: melody familiarization, detection of changes in the isolated melody, context familiarization, and finally, detection of changes in the melody in the three-voice context. Responses ("same" or "different"), when necessary, were made by pressing either *S* or *D* on a standard computer keyboard. Unlike Experiment 1, confidence ratings were not collected.

Isolated melody familiarity phase. During this phase participants were presented with a random ordering of two repetitions of the melody in each frequency range. Participants were informed that the overall frequency of the piece would vary across presentations but that the melody would be constant despite these changes in the overall frequency. During this phase participants were required only to sit quietly and listen to the stimuli for the purpose of familiarization. Participants were informed that their ability to detect single note changes in the melody would be tested in the next phase of the experiment.

Isolated melody test phase. During the melody test phase, changes occurred on half the trials and participants were required to indicate by keypress whether or not a change was present. As in Experiment 1, changes occurred with equal probability in three temporal locations (Notes 6, 8, or 10). During the melody test, as in the later test in the context of the accompaniment, changes consisted of a single note change. Participants received a total of 36 *different* trials (2 repetitions of 2 changes \times 3 change locations \times 3 melody frequency regions) and 36 *same* trials (12 repetitions of each of the 3 melody frequency regions). Presentation of the stimuli was randomized within these restrictions.

Context familiarity phase. Following a short break, participants were familiarized with the melody in the context of either a three-voice polyphonic or homophonic arrangement. As in the melody familiarity phase, no responses were required; participants simply sat and listened to the musical pieces. Participants were instructed to try to "hear out or perceptually isolate" the melody. Participants were further informed that the next task would be to detect changes in the melody while ignoring changes occurring elsewhere in the musical piece.

No changes were present during the familiarization phase. Participants were presented with two repetitions of each melody location within the given compositional style for that experimental session.

Detecting changes in context. After listening to the melody as part of the three-voice compositions, participants began the discrimination test session. Participants were told that every trial would contain a single note change and the task was to discriminate changes in the melody, which could occur in any voice, from changes in the accompaniment. Participants were further instructed to try to focus on the melody and ignore the other voices. Thus, the task in this phase was similar to that performed in the isolated melody test (i.e., respond "same" or "different" only to the melody).

On each test trial, participants were randomly presented with one of the three pieces (i.e., melody locations), with a single chord-related or chord-unrelated change occurring in one of the three voices and in one of three temporal locations (Note 6, 8, or 10). This resulted in a total of 18 "melody-different" stimuli and 36 "melody-same" stimuli. To keep the probability of a melody-same trial at .5, the melody-different trials were repeated twice for every repetition of the melody-same trials. Participants were presented with a total of 144 trials (72 melody-same and 72 melody-different trials, representing two repetitions of each melody-same trial and four repetitions of each melody-different trial).

Within this final phase of the experimental session, participants were first presented with four repetitions of the melody in isolation before beginning the "in-context" test. In addition, the original melody was presented in isolation after every seven test trials to refresh the participant's memory for the melody. As in Experiment 1, during the reminder trials a computer monitor displayed "*Original piece do not respond!*" to ensure that participants knew when the original melody was being presented.

Results and Discussion

As in Experiment 1, hit and false-alarm rates were calculated for each participant. In Experiment 2, false alarms were defined as responding that a change had occurred in the melody when in fact the change had occurred in the accompaniment. Hits were defined as responding accurately that a change had occurred in the melody. False alarm and hit probabilities were converted into Z scores and used to calculate d' . The d' scores were then entered into an ANOVA for further analysis. Separate ANOVAs were conducted for the detection of changes in the melody in isolation and in the three-voice context.

Melody in Isolation

The data collected across the two experimental sessions were entered into a two-way ANOVA with session (Day 1 vs. Day 2) and training (musician vs. nonmusician) entered as factors. This analysis revealed a significant main effect of training, $F(1, 10) = 9.07, p < .05$. Although both groups were performing at greater than 90% accuracy on the melody test during the second testing session, overall d' was significantly higher for musicians ($d' = 5.25, SE = .25$) than for nonmusicians ($d' = 4.30, SE = .43$). Neither the main effect of session nor the Session \times Training interaction was significant.

Identifying Melody Changes in Three-Voice Contexts

The last column in Table 1 presents a summary of the pattern of significant results obtained in the in-context conditions of the present experiment. Overall, the pattern of results is very similar to that observed by Acker and Pastore (1996b), who tested highly trained musicians with the same task and stimuli. A detailed analysis of the current results, provided in a following section, required a comparison with the results of Experiment 1.

Effects of Training

Within the three-voice context, musicians performed the detection/discrimination task significantly better than nonmusicians ($d' = 2.70$ and 1.95 , respectively). Figure 4 presents the Change Type \times Compositional Style \times Melody Location interaction collapsed across the levels of training (Panel A) and separately for musicians (Panel B) and nonmusicians (Panel C). As shown in Figure 4 and indicated by the fact that the effects of training did not interact with any other factor, the general pattern of performance for musicians and nonmusicians was very similar. Thus, musical training did not produce differences in the pattern of performance for either an integrative (Experiment 1) or selective (Experiment 2) listening task.

Comparison of Experiments 1 and 2

The d' data from Experiments 1 and 2 were entered into an ANOVA to directly compare performance across the integrative (Experiment 1) and the selective (Experiment 2) listening tasks. As was the case when each experiment was considered separately, musicians had significantly higher detection rates than did nonmusicians, $F(1, 20) = 82.07, p < .05$, but the level of musical

training did not interact with task (integrative vs. selective listening) or any of the other variables in the current study.

A significant main effect of compositional style was found, $F(1, 20) = 4.94, p < .05$, indicating that across the two tasks error detection was more difficult in the polyphonic context ($M = 2.18$) than in the homophonic context ($M = 2.56$). This main effect was qualified by a significant Task \times Compositional Style interaction, $F(1, 20) = 10.61, p < .05$, as well as a significant Task \times Compositional Style \times Melody Location interaction, $F(1, 40) = 6.66, p < .05$. A comparison of Figures 2 and 4 (the Melody Location \times Compositional Style \times Change Type interaction for each experiment) shows that the polyphonic composition with the melody in the high voice produced the lowest detection performance in the integrative listening task (Experiment 1) but produced the highest level of performance in the selective listening task (Experiment 2). Thus, the earlier conjecture that the presence of the asynchronous line in the middle voice increased segregation was supported. That is, this specific polyphonic piece was least compatible with performing an integrative listening task but was most compatible with the demands of the selective listening task.

An across-experiment comparison of performance in the homophonic context (top right panel in Figures 2 and 4) revealed that performance with the homophonic compositions was not as dramatically affected by the shift in task requirements. Performance decreased for all melody locations when the task shifted from integrative listening to selective listening, and this decrease was most apparent for the homophonic composition with the melody in the middle voice. One possible interpretation of this finding is that listeners, especially in the vertically integrated homophonic compositions, easily and quickly detected an error on the basis of dissonance resulting from integration across voices. This same integration, however, limited their ability to then determine the voice in which the error occurred. This conjecture fits with Sloboda's (1985, p. 170) proposal, as well as with data provided by Acker and Pastore (1996c) for chord mistunings, that listeners can often detect an error without being able to specify the voice in which it occurred, especially when the mistuning is in the middle voice.

General Discussion

The current experiments were designed to evaluate the relative influence of the two types of auditory processing, primary (stimulus-driven) and secondary (scheme-driven), on the ability to detect changes to three-voice musical pieces. The role of primary processes was addressed by comparing change-detection performance as a function of compositional style (i.e., polyphonic or homophonic). The current stimuli were constructed to be distinct examples of polyphony and homophony, differing along dimensions previously shown to influence perceptual grouping. On the sole basis of the predicted effects of stimulus-driven processes, it was expected that the polyphonic compositions, defined by asynchrony between voices, would result in more segregation between voices. The homophonic compositions, having synchronous onset/offset of notes across voices, were expected to produce a higher degree of vertical integration. The operation of secondary (or scheme-driven) processes was addressed in two ways. First, two groups of listeners (musicians and nonmusicians) who were assumed to differ in the degree or strength of their music schemes

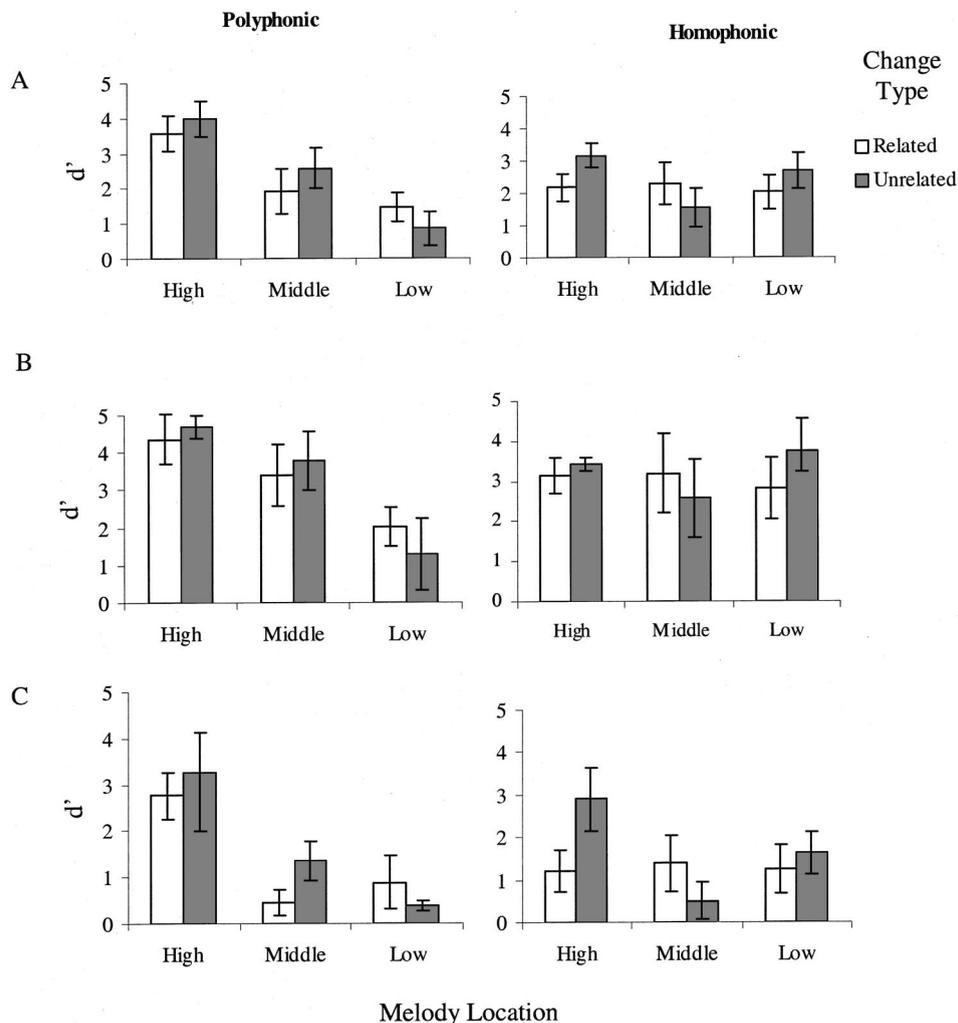


Figure 4. Sensitivity to the presence of a single note change in Experiment 2 as a function of change type (related and unrelated), compositional style (polyphonic and homophonic), and melody location (high, middle, and low voice) averaged across all listeners (Panel A) and separately for musicians (Panel B) and nonmusicians (Panel C).

were compared. Second, attentional flexibility, or the ability to use strategic listening as a function of task demands, was also tested under both an integrative and a focused attention task.

Implications for Scene Analysis

The results of the present study indicate that perceptual grouping is largely determined by properties of the physical stimulus and thus by processes operating at the primitive level. Secondary processes, as characterized by the differences observed between musicians and nonmusicians, appear to increase the listener's ability to use information provided by the output of the primary processes but do not appear to lead to a perceptual reorganization. It is possible that the demands of the current task emphasize primitive processes or features of the stimuli that are readily available to both groups of listeners. For example, with low memory load tasks, primitive processes may be of greater importance, with scheme-driven processes playing no more than a sup-

portive role, perhaps by enhancing the signal-to-noise ratio and increasing consistency in grouping. Schemes, rather than reorganizing incoming information, may play a greater role in the maintenance and retrieval of information. Thus, changing the nature of the task to rely more heavily on stored representations may place greater weight on scheme-driven processes. Although the current task did have a memory component, listeners only had to remember the standard over a maximum of seven trials and a minimum of one trial. A task that requires a longer retention period may provide different conclusions as to the role of schemes.

Musical Training and Scheme-Driven Processing

In both of the present experiments, musicians were consistently better at detecting a single diatonic change than were nonmusicians, but the overall pattern of performance did not differ significantly for the two groups in either of the two tasks (integrative and focused listening). Thus, the present study, using musically valid

stimuli, lead us to the same conclusion drawn by Jones et al. (1995). Specifically, musical training does not appear to lead to a greater attentional flexibility or greater freedom from the output of stimulus-driven processes. Since musicians show a consistent advantage over nonmusicians when tested with musical stimuli, there can be no doubt that musical training enhances performance. Despite this overall advantage, the performance of musically trained listeners was no less dependent on stimulus attributes than was that of nonmusicians. That is, the relative ability to detect and discriminate a single note change as a function of compositional style was the same for musicians and nonmusicians. It thus appears that performance in the present study reflects the operation of processes shared by both musician and nonmusicians. These shared processes could reflect either primitive grouping mechanisms or basic schemes that are either based on perceptually salient stimulus relationships or are developed through relatively passive exposure to Western music.

A growing body of literature indicates that sensitivity to changes or “mistakes” in relatively complex, culturally specific music develops without any explicit musical training (Lynch, Eilers, Oller, Urbano, & Wilson, 1991; Trainor & Trehub, 1992, 1993, 1994). Even adults who claim to be tone deaf show a pattern of sensitivity to the tonal organization of melodies similar to that produced by musically trained participants (Mawhinney, 1987, as cited in Cuddy, 1993). Thus, it appears that some implicit knowledge of the rules underlying the listener’s culturally specific music system, as taught through formal training, develops early in life and is retained through adulthood. In the present study, the lack of qualitative differences in performance between musicians and nonmusicians when responding to relatively subtle changes (i.e., all changes were diatonic) within a relatively complex context indicates that nonmusicians, like musicians, are very sensitive to musical form. These findings are not surprising because musical schemes likely developed to take advantage of at least a subset of primitive perceptual organization processes.

References

- Acker, B. E., & Pastore, R. E. (1996a, May). *Directed attention and perception of frequency changes*. Paper presented at the 131st Meeting of the Acoustical Society of America, Indianapolis, IN.
- Acker, B. E., & Pastore, R. E. (1996b). Melody perception in homophonic and polyphonic contexts. In B. Pennycook & E. Costa-Giomi (Eds.), *Proceedings of the Fourth International Conference on Music Perception and Cognition* (pp. 453–458). Montreal, Quebec, Canada: McGill University, Faculty of Music.
- Acker, B. E., & Pastore, R. E. (1996c). Perceptual integrality of major chord components. *Perception & Psychophysics*, *58*, 748–761.
- Barsz, K. (1988). Auditory pattern perception: The effect of tonal frequency range on the perception of temporal order. *Perception & Psychophysics*, *43*, 293–303.
- Beauvois, M. W. (1998). The effect of tone duration on auditory stream formation. *Perception & Psychophysics*, *60*, 852–861.
- Botte, M., Drake, C., Brochard, R., & McAdams, S. (1997). Perceptual attenuation of nonfocused streams. *Perception & Psychophysics*, *59*, 419–425.
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*, 244–249.
- Cuddy L. (1993). Melody comprehension and tonal structure. In T. J. Tighe & W. J. Dowling (Eds.), *Psychology and music* (pp. 19–38). Hillsdale, NJ: Erlbaum.
- Dai, H., Scharf, B., & Buus, S. (1991). Effective attenuation of signals in noise under focused attention. *Journal of the Acoustical Society of America*, *89*, 2837–2842.
- Dewitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals. *Perception & Psychophysics*, *41*, 73–84.
- Encore. (1995). (Version 4.04) [Computer software]. Half Moon Bay, CA: Passport.
- Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 293–307.
- Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal function within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 579–594.
- Lackner, J. R., & Goldstein, L. M. (1974). Primary auditory stream segregation of repeated consonant-vowel sequences. *Journal of the Acoustical Society of America*, *56*, 1651–1652.
- Lynch, M. P., Eilers, R. E., Oller, K. D., Urbano, R. C., & Wilson, P. (1991). Influences of acculturation and musical sophistication on perception of musical interval patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 967–975.
- Macmillian N. A., & Creelman C. D. (1991). *Detection theory: A user’s guide*. New York: Cambridge University Press.
- Palmer, C., & Holleran, S. (1994). Harmonic, melodic, and frequency height changes in the perception of multivoice music. *Perception & Psychophysics*, *56*, 301–312.
- Pastore, R. E., & Scheirer, C. J. (1974). Signal detection theory: Considerations for general application. *Psychological Bulletin*, *81*, 945–958.
- Sloboda, J. A. (1985). *The musical mind*. Oxford, England: Clarendon Press.
- Sloboda, J. A., & Edworthy, J. (1981). Attending to two melodies at once: The effect of key relatedness. *Psychology of Music*, *9*, 39–43.
- Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants’ and adults’ sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 394–402.
- Trainor, L. J., & Trehub, S. E. (1993). Musical context effects in infants and adults: Key distance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 615–626.
- Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*, 125–132.

(Appendix follows)

Appendix

Six Compositions Used in the Experiments

Homophonic

High Melody

A musical score for a homophonic composition with a high melody. It consists of three staves: a vocal line in the treble clef with a high range of notes, and two piano accompaniment staves (treble and bass clefs) in the lower registers. The music is in 4/4 time and features a simple, steady harmonic accompaniment.

Middle Melody

A musical score for a homophonic composition with a middle melody. It consists of three staves: a vocal line in the treble clef with a middle range of notes, and two piano accompaniment staves (treble and bass clefs) in the lower registers. The music is in 4/4 time and features a simple, steady harmonic accompaniment.

Low Melody

A musical score for a homophonic composition with a low melody. It consists of three staves: a vocal line in the treble clef with a low range of notes, and two piano accompaniment staves (treble and bass clefs) in the lower registers. The music is in 4/4 time and features a simple, steady harmonic accompaniment.

Polyphonic

High Melody

A musical score for a polyphonic composition with a high melody. It consists of three staves: a vocal line in the treble clef with a high range of notes, and two piano accompaniment staves (treble and bass clefs) in the lower registers. The music is in 4/4 time and features a more complex, rhythmic accompaniment.

Middle Melody

A musical score for a polyphonic composition with a middle melody. It consists of three staves: a vocal line in the treble clef with a middle range of notes, and two piano accompaniment staves (treble and bass clefs) in the lower registers. The music is in 4/4 time and features a more complex, rhythmic accompaniment.

Low Melody

A musical score for a polyphonic composition with a low melody. It consists of three staves: a vocal line in the treble clef with a low range of notes, and two piano accompaniment staves (treble and bass clefs) in the lower registers. The music is in 4/4 time and features a more complex, rhythmic accompaniment.

Note. The compositions differed in terms of the synchrony of the voices (homophony and polyphony) and the location of the melody (high, middle, and low).