Perception of Mode, Rhythm, and Contour in Unfamiliar Melodies: Effects of Age and Experience

Author(s): Andrea R. Halpern, James C. Bartlett and W. Jay Dowling

Source: *Music Perception: An Interdisciplinary Journal*, Vol. 15, No. 4 (Summer, 1998), pp. 335-355
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40300862
Accessed: 09/04/2013 20:46

Music Perception
Summer 1998, Vol. 15, No. 4, 335–355

Perception of Mode, Rhythm, and Contour in Unfamiliar Melodies: Effects of Age and Experience

ANDREA R. HALPERN
Bucknell University

JAMES C. BARTLETT & W. JAY DOWLING
University of Texas at Dallas

We explored the ability of older (60–80 years old) and younger (18–23 years old) musicians and nonmusicians to judge the similarity of transposed melodies varying on rhythm, mode, and/or contour (Experiment 1) and to discriminate among melodies differing only in rhythm, mode, or contour (Experiment 2). Similarity ratings did not vary greatly among groups, with tunes differing only by mode being rated as most similar. In the same/different discrimination task, musicians performed better than nonmusicians, but we found no age differences. We also found that discrimination of major from minor tunes was difficult for everyone, even for musicians. Mode is apparently a subtle dimension in music, despite its deliberate use in composition and despite people's ability to label minor as "sad" and major as "happy."

When composers use compositional techniques like theme-and-variations, rondo, or even simple repeats, they are obviously counting on us to remember some aspects of music after just one or a few hearings. Several studies have shown that musicians are capable of correctly classifying musical phrases as belonging to a particular piece of music (Pollard-Gott, 1983), even in a relatively unfamiliar musical idiom (Krumhansl, 1991). Nonmusicians are less able to do this in a sophisticated way, but Pollard-Gott found that even nonmusicians could classify passages of a Liszt sonata using basic musical dimensions such as pitch range and loudness. In fact, Welker (1982) showed that nonmusicians can abstract commonalities among variations generated from a theme well enough that they will falsely accept the nonpresented theme as having in fact been presented.

Address correspondence to Andrea R. Halpern, Psychology Department, Bucknell University, Lewisburg, PA 17837. (e-mail: ahalpern@bucknell.edu)

Smith (1997) points out that the field of music perception would do well to pay more attention to the cognitive capabilities of musically inexperienced people. Much research indicates that novices fail to appreciate many of music's regularities, particularly with regard to tonal organization. Yet, novices clearly appreciate and seek out opportunities to listen to music. This paradox, Smith points out, needs to be researched by studying nonmusicians in music cognition research. What do nonmusicians hear and understand about a melody after a brief exposure?

One attempt to answer that question was a study by Halpern (1984). In that study, young adult musicians and nonmusicians heard melodies that were generated from one of two themes. The variations differed in specific ways from the theme. The variations could differ in rhythm (which also changed the meter), contour (the theme was written in retrograde, which inverted the contour), or mode (major and minor). Figure 1 shows half of the stimulus set where one note pattern generated seven other related melodies that could differ from the original pattern by one, two, or all three musical dimensions.

In the first task, participants were asked to rate the similarity of each tune to every other tune, after transposition between the keys of C and F to avoid comparisons based strictly on absolute pitch values. A clustering analysis showed that nonmusicians returned a very orderly set of similarity relations that roughly reflected the factorial nature of the stimulus set (the clustering solution accounting for 0.80 of the variance in the data). This occurred despite their subjective impression that "all the melodies sounded alike." Melodies differing by rhythm or contour were seen as being more distinct than those differing by mode. The musicians returned roughly the

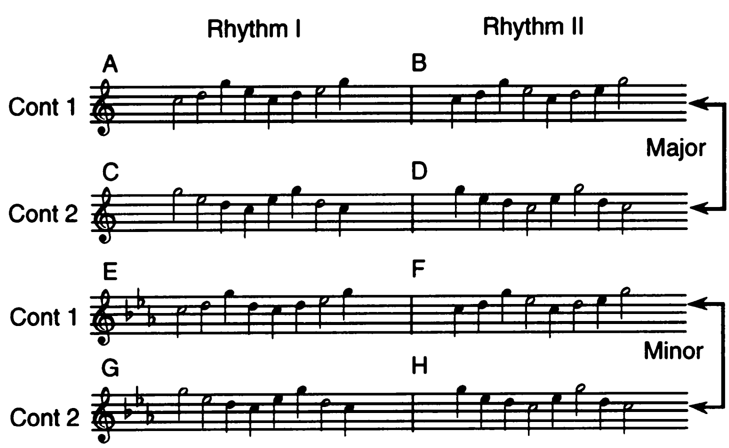


Fig. 1. Stimulus melodies used in Experiments 1 and 2 here, which were a subset of those used in Halpern (1984).

same clustering solution, although the fit of the model to their data was somewhat less satisfactory (accounting for 0.60 of the variance in the data). This was presumably due to the musicians being more sensitive to the multiple ways that the melodies could be grouped.

In a second experiment, the melodies in Figure 1 were presented in a learning task. A new set of participants were asked to learn arbitrary letter names for each melody during a training phase. In a test phase, naming errors constituted a confusion matrix, which was again subjected to a clustering analysis. Although musicians performed at a higher level than nonmusicians (54% correct vs. 31%; chance = 12.5%), both groups once again returned an orderly pattern of results that matched the similarity ratings of Experiment 1 quite well. Nonmusicians confused major/minor pairs most often and confused different-contour and rhythm pairs least often. Musicians also confused different-contour pairs least often but, surprisingly, made some errors on different-rhythm pairs as well as on major/minor pairs. Overall, the correspondence between the "off-line" task of similarity judgment and the more "on-line" learning task was reassuringly close; items judged as more similar are apparently confusable in memory in a predictable way.

From this study, Halpern (1984) concluded that even nonmusicians can hear structures in unfamiliar music, although the distinction between major and minor seemed unclear to them. The ordering of salience of these dimensions corresponds to results found from many other studies in the literature, using both musician and nonmusician samples. Rhythm and contour are clearly salient dimensions in music. For instance, both Dowling (1973) and Jones and Ralston (1991) found that presenting an old melody with a new rhythm significantly impairs the ability to recognize the melody. Contour has also been shown to be an important organizing feature for remembering new melodies, especially over short, unfilled intervals (Dowling, 1991; Dowling, Kwak, & Andrews, 1995).

The ability to distinguish major from minor has been less studied, especially with regard to mode in whole melodies. What findings we do have are somewhat inconsistent about the salience of this dimension. On the one hand, mode is certainly used in Western music as a distinct compositional device, and evidence suggests that even untrained listeners and children can recognize the archetypal attribution of "happy" to major scales and "sad" to minor scales. As an example, Gerardi and Gerken (1995) played unfamiliar melodies in major and minor modes to 5-year-olds, 8-year-olds, and college students unselected for musical background. Listeners simply had to pick a happy or sad label for each one (the children pointed to happy and sad faces). The youngest children did not distinguish the melodies by affect, but the older children and college students reliably rated the major melodies as happier than the minor melodies.

On the other hand, when not explicitly asked to associate melodies with affect, even adults have a difficult time distinguishing major from minor mode melodies. Recall that Halpern (1984) found that not only did tunes identical except for mode elicit the highest similarity ratings, nonmusicians and to some extent musicians most often erred in the learning task by giving the label of a tune's major or minor version to a target tune. In a different kind of memory test, Madsen and Staum (1983) presented nonmusic majors (degree of musical training unreported) with a target melody on each trial, followed by eight other melodies. One was the identical melody again (untransposed), six were unrelated melodies serving as interference, and another melody was the same as the original except for mode or rhythm changes. Madsen and Staum found that when an error was made in identifying the second occurrence of the original melody, the most common error was choosing the same-except-for-mode melody and the next most common was the rhythmic variation. Although these data were presented in the form of rank orders rather than in a rigorous quantitative analysis, results of both this and the Halpern study are consistent with the hypothesis that many people cannot easily distinguish major from minor melodies under challenging memory conditions.

The current study was an attempt to replicate the Halpern (1984) results with respect to similarity ratings and to add a very direct test of whether people can reliably distinguish major and minor pairs. From our previous results, we could not tell whether major and minor are confused in memory only after time and interference have transpired, or whether the lowered third that characterizes minor from major is not even perceptually salient. To that end, we devised a same/different discrimination task where the "different" melody pairs differed by either rhythm, contour, or mode. Although the melodies within a pair were in different keys, the interval between the pairs was only 4 s and was silent (no interference). The literature on affect and mode would suggest that even untrained listeners should be able to perform this discrimination task, but the memory studies cited earlier lead to the hypothesis that discriminating major from minor should not only be more difficult than rhythm and contour discriminations but may be nearly impossible.

This study also examined two listener variables. One was musical experience. Halpern (1984) found some differences between musicians and nonmusicians with respect to the ordering of salience of musical dimensions, although the similarities were more striking than the differences. In light of the points that Smith (1997) brings up about the relative absence of cognitive musical structures among nonmusicians, it seemed worthwhile to consider this variable again. At the other end of the competence spectrum, we wanted to see if the difficulty with major/minor decisions would extend to musicians. Despite the implicit and explicit knowledge that musicians have of this distinction, would they be impaired on an absolute or relative basis with this particular discrimination?

The second listener variable we examined here was age. In each of our two experiments, we tested healthy younger adults (college students) and older adults (60–80 years old). Very little work has examined the relationship between adult aging and musical cognition, most of that work being from our laboratory. Our field has tended to assume that young adults represent an end-stage in musical development. But this assumption must be examined if we wish to know whether adults' understanding of music cognition continues to mature, as a lifetime of musical listening experience accrues, or perhaps declines, as certain biological processes become less efficient.

In our examination of aging and musical cognition, we have found some tasks to be more sensitive to age than others. As in other domains, older adults show impaired old/new recognition relative to younger adults for both familiar and unfamiliar tunes over retention intervals from a few minutes to an hour (Bartlett, Halpern, & Dowling, 1995; Halpern, Bartlett, & Dowling, 1995). We also investigated age effects in a transposition detection task in which an unfamiliar tune was presented four times in different keys, followed by 5 s of silence, and then a comparison tune (Halpern et al., 1995). The comparison tune was either transposed exactly, or was transposed with a change of contour, or with the same contour but with two intervals changed. We found significant age-related impairments in most of the four experiments in that paper, but mainly for the discrimination between exact transposition and changed contour trials (which was the easier of the two discriminations). For these same tasks, we found little or no effect of musical experience.

In contrast, we have found some tasks that are relatively impervious to age effects and more susceptible to the influence of musical experience. Halpern et al. (1995) found that the discrimination between exact transpositions and same-contour lures showed few age effects. Similarly, Halpern, Kwak, Bartlett, and Dowling (1996) showed that older adults were not impaired (and in one analysis, were superior) relative to younger adults in showing their knowledge of the tonal hierarchy by the use of the probe-tone method (Krumhansl & Shepard, 1979). In both of these tasks, experience had larger effects than age.

We tentatively concluded from this series of experiments that some musical tasks, such as abstracting contour or remembering a series of tunes, require use of general purpose perceptual or memory skills, which are known to show age-related impairments in other domains. Other tasks, such as detecting the exact intervals in a transposed tune, or developing a sense of the tonal hierarchy, make use of more domain-specific skills, which once developed, do not seem to decline with age and may even increase.

With this dichotomy in mind, we wondered whether age would change the way in which musical dimensions were abstracted from a piece of music. The similarity rating task resembles the probe-tone task in that they are both designed to capture the "semantic" knowledge of music, in contrast to a memory task imposing more stringent cognitive challenges to on-line processing. Because we found no age-related decline in the probe-tone task (Halpern et al., 1996), we predicted older adults would show a similar pattern to younger adults in their reaction to which musical dimensions make a tune more or less resemble another. This finding would be consistent with Charness's findings from the domain of chess (summarized in Charness, 1989). He found that older chess players were less successful than younger ones of equivalent skill on a chess memory task, but were as good as younger players in the more reflective task of choosing the best next move from a given chess position.

Following this logic, age effects in the discrimination task (requiring use of a more on-line skill) were hypothesized to follow the dichotomy just outlined. Rhythmic and contour discrimination can be considered general perceptual skills, as people need to be able to distinguish contours in speech patterns and rhythms in movement and visual patterns, as well as in music. Thus, we might expect age effects in a same/different task tapping those dimensions, as Halpern et al. (1995) showed that age effects were most prominent in detection of contour violation (despite this being a relatively easy task). In contrast, we may think of mode discrimination as being more specifically musical. It is hard to think of any other area in which the distinction of a half step in the third position of the scale changes the "meaning" of a pattern as mode does in music. This type of discrimination is most similar to the exact/same contour discrimination in Halpern et al. (1995), in that the comparison sequence has only one pitch class changed from the initial sequence (compare Melody A with Melody E in Figure 1). It was this comparison that showed minimal age effects but substantial experience effects. Thus, despite the overall difficulty we expected people to have in distinguishing major from minor, this task would conceivably show fewer age differences than detecting rhythm and contour change.

Finally, having two different age and experience groups allows us to look at the interaction of those factors. In previous work, we have found little evidence that increased experience can "compensate" for age-related impairments. However, perhaps extraction of regularities heard in novel music is stable or accrues with age. Thus in the current experiment, we wondered if increased experience and increased age together would lead to an increased ability to extract dimensions from music. If so, we would expect our older musicians to generate patterns of data distinct from the other groups, perhaps in the form of increased regularity in similarity ratings or in their ability to hear the major/minor distinction more acutely.

Experiment 1

METHOD

Subjects

Young subjects consisted of 24 volunteer undergraduate psychology students at Bucknell University. Of these, 12 musicians had received an average of 11.75 years of private music lessons (mean age = 18.83 years). The 12 nonmusicians had received an average of 0.63 years of music lessons (mean age = 18.50 years). Older subjects lived in the Dallas/Ft. Worth area. Those classified as musicians consisted of 12 senior citizens with an average of 13.9 years of formal lessons (mean age = 69.64 years). Many older adults who were currently or had been active performers were at least partly self-taught, and so this figure underestimates their years of musical experience. Older subjects classified as nonmusicians consisted of 12 senior citizens with an average of 0.58 years of music lessons (mean age = 71.58 years).

Materials

Eight melodies composed by Halpern (1984) were used (see Figure 1). All eight melodies were generated from Melody A. Melody A was transposed to the minor (Melody E). Melodies A and E were then altered rhythmically (Melodies B and F, respectively). These melodies were then written in retrograde (Melodies C, D, G, and H), which inverted the contour. Thus eight melodies were created as a factorial combination of the two modes, contours, and rhythms.

Melodies were played on a Yamaha PSR-500 synthesizer using a piano voice and recorded digitally on a Zenith Supersport computer via the Cakewalk MIDI sequencer. Tempo was set at 120 beats per minute, and all notes were equalized in duration and intensity. Transposed and altered versions were created directly on the sequencer and recorded onto audiotape via a Marantz stereo cassette recorder.

Two tapes were prepared, each consisting of 28 pairs of the eight melodies in random order. Pair members were separated by a 4-s pause. Pairs were separated by a 6-s pause. Any pair presented in order AB on one tape was presented as BA on the other tape. In each pair, one melody was played in the key of C major or minor and the other in the key of F major or minor. This was done to ensure that similarity judgments could be made independently of key. The key of each melody and the key of the first melody of the pair were counterbalanced over the sequences. Tapes were played back on the Marantz cassette recorder through Acoustic Research stereo speakers.

Procedure

Listeners were tested individually. Each person was first given a musical background questionnaire, followed by a 20-item vocabulary test taken from the second half of the Wechsler Adult Intelligence Scale. In order to familiarize listeners with the stimulus materials, all eight melodies were played once. The listeners then heard one of the two tapes; half of the participants heard each counterbalancing tape. The task was to make similarity judgments on each pair of melodies. A scale of 1 to 7 was used, where 1 indicated that the melodies were not very similar and a 7 indicated that they were very similar. This scale was printed on the answer sheet and available for consultation throughout the session, which lasted about 30 minutes. After the procedure, subjects were debriefed.

RESULTS

Vocabulary

As is typical, older listeners outscored younger listeners on the vocabulary measure. Where the maximum score equaled 40, older people scored an average of 27.2 points and younger people an average of 21.1 points, $F(1, 42) = 6.52, p = .01$). Musicians and nonmusicians did not differ (means = 23.1 and 24.3), nor did age and experience interact.

Clustering

A mean similarity score for each pair of melodies (ignoring order of melodies within a pair) was calculated across listeners for each group. The means were then analyzed by way of a clustering program called ADDTREE (Sattath & Tversky, 1977). ADDTREE analyzes similarity data and represents the proximity of melodies to each other as a tree with a vertical trunk and horizontal branches (Figures 2 and 3). The melodies that correspond to those in Figure 1 are shown at the far right of the tree. Similarity between melodies is represented by the sum of the branches connecting two melodies, and the length of a branch out to a cluster is representative of the cluster's distinctiveness. ADDTREE also calculates an r^2 value that represents percentage of variance in the data accounted for by the model.

The ADDTREE solutions accounted for a large proportion of the variance in each group's data (r^2 = 0.90, 0.81, 0.82, and 0.79 in younger nonmusicians, younger musicians, older nonmusicians, and older musicians, respectively). Visual inspection shows clearly that each group reflected the factorial nature of the stimulus set in their similarity ratings.

The trees for the younger musicians and nonmusicians were nearly identical, and indeed their ratings for each pair correlated at $r(26) = 0.87$. The first branching strongly grouped all melodies of Rhythm I and all melodies of Rhythm II, showing that same-rhythm melodies were considered to be similar whereas melodies of differing rhythm were less likely to be considered similar. Next, within each rhythm group, melodies of the same contour were grouped together. This suggests that for a given rhythm, melodies that shared a contour were considered to be similar whereas melodies that differed by contour were considered to be distinctive. The final group consists of the major/minor pairs. Melodies that differed only by mode were considered the most similar.

The trees for older participants are quite similar to this pattern, but differ in a few aspects, and the older musicians and nonmusicians differ some-

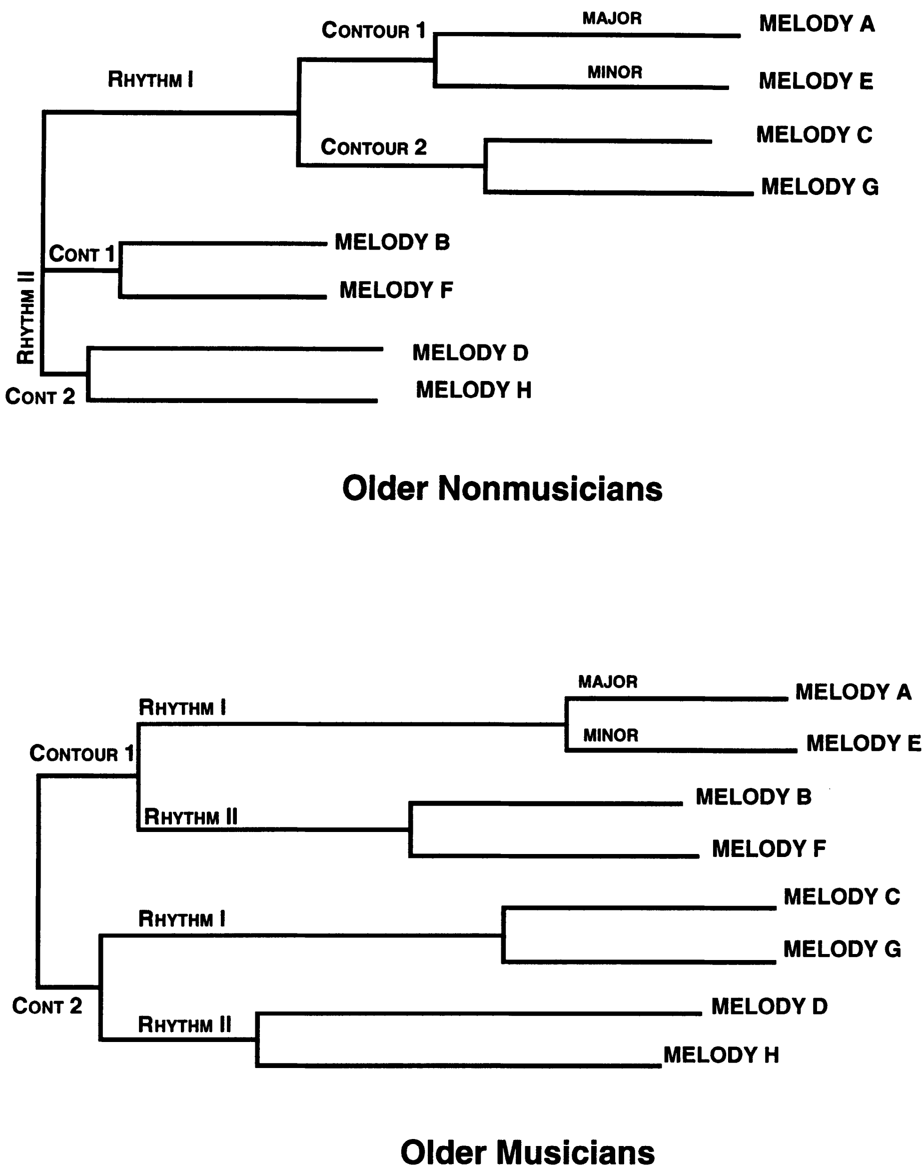
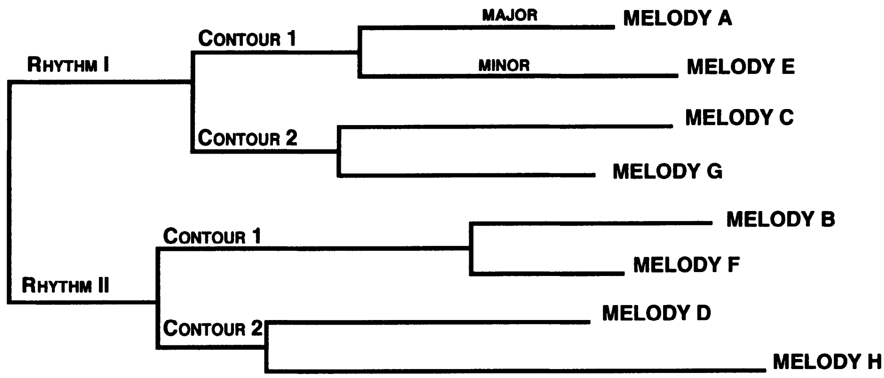
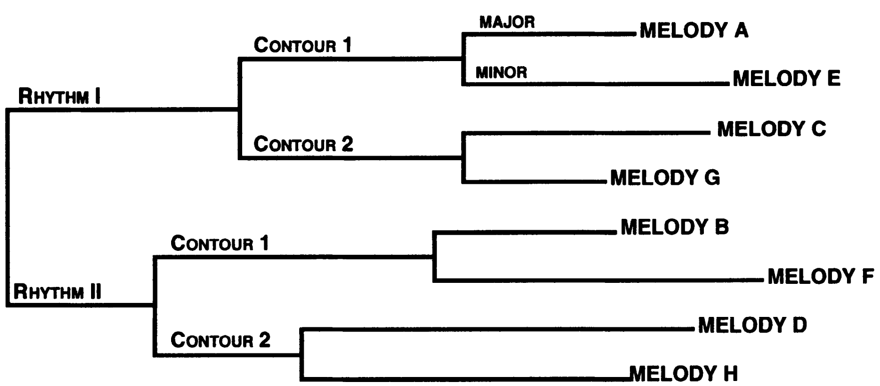


Fig. 2. ADDTREE solution for older musicians and nonmusicians in Experiment 1. Variance accounted for by the model = 0.79 and 0.82, respectively.

what from each other, $r(26) = 0.68$. The results from the older nonmusicians resemble the results from younger groups quite a bit: Rhythm was the most distinctive dimension and mode the least distinctive, $r(26) = 0.76$ and 0.73 with the younger musicians and nonmusicians, respectively. However, compared with the younger groups, rhythm was not as strong an organizing factor, seen by the fact that Rhythm II melodies "attach" directly onto the



Younger Nonmusicians



Younger Musicians

Fig. 3. ADDTREE solution for younger musicians and nonmusicians in Experiment 1. Variance accounted for by the model = 0.81 and 0.92, respectively.

main branch, instead of forming a distinctive cluster. Rhythm I melodies were, however, strongly grouped.

For the older musicians, the most distinctive factor was contour. That is, melodies that differed by contour were considered the least similar. Within each contour group, they next grouped melodies by rhythm. However, note that the rhythm grouping was quite distinctive (long branches) whereas the contour group was less distinctive, which resembles the other trees. As with the younger subjects, the final group was mode, suggesting that melodies that differed only by mode were considered to be the most similar. The older musicians' ratings correlated positively with those of younger musicians and nonmusicians, $r(26) = 0.74$ and 0.70, respectively.

Similarity and Shared Dimensions

As another way of looking at the data, we divided the stimulus items into those that differed on one, two, or three of the musical factors. For instance, Melody A differs from Melodies B, E, and C by one factor (rhythm, contour, and mode, respectively); it differs from Melodies D, F, and G by two factors (rhythm and contour, rhythm and mode, contour and mode, respectively) and from Melody H by all three factors. In this stimulus set, there are 12 one- and two-factor pairs, and 4 three-factor pairs. We tested whether pairs differing on one factor would be rated as more similar than those differing on two factors, and likewise compared with those differing on all three factors and whether this pattern would vary for the different listener groups.

An analysis of variance used two between-subjects factors (age and experience) and one within-subjects factor (stimulus type). We found a substantial main effect of stimulus type in the expected direction, $F(2, 88) = 83.27$, $p < .001$. Items differing by only one factor elicited an average rating of 4.10 (where 7 is the highest similarity score), compared with ratings of 3.38 and 2.52 for two-factor and three-factor items, respectively. This pattern did not vary as a function of age or experience, as none of the interactions of item type with age or experience were statistically significant. Nor do the pattern of means reveal any trend towards a difference. We did find an unexpected interaction of age and experience, $F(1, 44) = 5.09$, $p = .03$, although effects not involving stimulus type simply reflect the tendency of the different groups to use higher or lower similarity ratings and are thus not terribly germane to our main point. The interaction was in the form of a crossover, where among younger listeners, nonmusicians used lower similarity scores overall than musicians, but among older listeners, nonmusicians used higher similarity scores than musicians.

DISCUSSION

Generally speaking, results from Experiment 1 replicated those from Halpern (1984). An exact comparison is not possible because the comparable experiment in the earlier article used 16 melodies in the similarity rating experiment rather than the 8 used here, where the additional 8 melodies were generated from a different note pattern (and therefore added a fourth musical dimension, that of note pattern). However, as before, the factorial nature of the stimulus set is reflected quite closely in the clustering results from everyone. As the proportion of variance accounted for was high in all groups, it seems that the similarity relationships were being responded to in a consistent and orderly way by all our listeners.

Specifically, we can see that three of the four listener groups here divided the stimulus set first by rhythm and then by contour, comparable to the most common pattern in the earlier findings. Although the older musicians reversed these two dimensions in their ordering of salience, their contour groupings were not as distinctive as their rhythm groupings. Also, older nonmusicians did not group Rhythm II melodies very strongly. Although only suggested by this qualitative analysis, older listeners may give more equal weighting to rhythm and contour than do younger listeners, who seem to divide the tunes more strongly on rhythm.

All four groups agreed that the major/minor pairs were the most similar pairs, also comparable to the results from Halpern (1984). This result cannot tell us if the listeners were hearing similarity between identifiably different melodies or were unable to tell the pair members apart. Experiment 2 was designed to answer that question.

From this experiment, we have only slight evidence that musicianship or age was affecting the similarity relationships heard within this set of tunes. Nonmusicians were quite able to respond to the regularities in this set, despite the transposition of a musical fourth or a fifth between the pair members in a trial. Age does not seem to diminish the ability to hear these regularities. Our quantitative analysis comparing trials in which tunes differed by one, two, or all three dimensions was consistent with this conclusion, as it revealed no differences in the way that musicians vs. nonmusicians or older vs. younger people rated the pairs. Also, the correlations of ratings among all the groups were consistently high and positive.

Experiment 2 used a same/different discrimination task to help determine whether similarity ratings would translate directly into discrimination ability. If so, then pairs differing only on rhythm should be the easiest to distinguish, followed by contour pairs. Pairs differing only on mode should be the most difficult to distinguish. If the reversal in ordering of salience of contour and rhythm for the older musicians really reflects a different way of processing the melodies, then their discrimination results should reflect that reversed ordering. We were also interested to see if age would have an overall detrimental effect on the discrimination task, or a specific effect on particular discriminations, as outlined in our opening remarks.

Experiment 2

In this experiment, we presented pairs of melodies that were exact transpositions of each other or were a transposition plus a mode, contour, or rhythm change. The task was to indicate whether the melodies were the same or different, using a 6-point response scale.

METHOD

Subjects

Young subjects consisted of 31 volunteer undergraduate psychology students at Bucknell University. Of these, 14 musicians had received an average of 14 years of private music lessons (mean age = 19.00 years). The 17 nonmusicians had received an average of 1.3 years of music lessons (mean age = 19.00 years). The 24 older listeners lived in the Dallas/Ft. Worth area. Those classified as musicians consisted of 12 senior citizens, with an average of 8.5 years of music lessons (mean age = 67.58 years). As before, it should be noted that some older adults who were active performers were nevertheless self or mostly self-taught, and so this figure underestimates their years of musical experience. Older subjects classified as nonmusicians consisted of 12 senior citizens with an average of 0.5 years of training (mean age = 68.42 years).

Materials

The eight melodies from Experiment 1 served as stimuli. Each of two counterbalancing tapes consisted of 16 Same trials and 24 Different trials. The Same trials paired each melody with itself, twice. The two instances of a Same trial for a particular melody differed only in whether the first member of the pair was in the key of C or F. The Different trials were of three types: Pairs could differ by Rhythm only (e.g., Melodies A and B), Contour only (Melodies A and C) or Mode only (Melodies A and E). Although there were four such pairs for each musical dimension, each item was repeated on the tape by exchanging the order and the key of each member of the pair. For instance, on Tape 1, one Mode trial consisted of Melody A in the key of C major paired with Melody E in the key of F minor. Later on the tape, Melody E was presented in C minor followed by Melody A in the key of F major. On the second tape, items were presented in a different random order, and for Different pairs, the order of items was reversed.

Tapes were prepared and played back in the same way as in Experiment 1, with the exception that an IBM 286 computer controlled the Cakewalk software. As before, items within a pair were separated by 4 s, and pairs were separated by 6 s.

Procedure

Participants were tested individually or in small groups. Sessions began as in Experiment 1, with a musical background questionnaire and the vocabulary items from the Wechsler Adult Intelligence Scale. Instructions then advised participants that they were going to hear 40 pairs of short, unfamiliar tunes. The first member of the pair would start on a certain note, and then a second melody would begin on a different note. The task was to rate confidence on a scale of 1 to 6 that the tunes were the Same or Different, except for the starting note (the word "transposition" was used with the musicians). Scale value 1 meant "sure different" and scale value 6 meant "sure same." The scale was in view at all times during the experiment.

Next, participants heard three practice trials with tunes not heard in the experiment. One was a Same trial and two were Different trials. If the practice trials were not answered correctly, the experimenter reviewed the correct answers and replayed the practice trials until they were all answered correctly. After the practice trials, one of the two counterbalancing tapes was played. If the 6-s response period was insufficient, the experimenter paused the tape briefly. The session, including debriefing, lasted about 30 min.

RESULTS

Vocabulary

Reliable vocabulary scores were available for 50 of the 55 participants (protocols for two young adults were lost; one older adult was not a native speaker of English, and two other older adults scored so low on the test, 4 and 6 out of 40, respectively, that it is likely they misunderstood the instructions). Once again, older listeners outscored younger listeners on the vocabulary measure. Older people scored an average of 29.4 points and younger people an average of 21.5 points, $F(1, 46) = 25.75$, $p < .001$. Musicians and nonmusicians did not differ (means = 26.3 and 24.6), but age and experience interacted, $F(1, 46) = 5.26$, $p = .03$. Among younger people, musicians (mean = 20.5) and nonmusicians (mean = 22.4) did not differ; among older people, musicians (mean = 32.1) exceeded nonmusicians (mean = 26.8).

Area Under the Memory Operating Characteristic Curve

Our main dependent measure was the area under the memory operating characteristic curve for discrimination between Same pairs and pairs that differed by Rhythm, Mode, or Contour. These area scores were computed using the confidence levels to provide an unbiased estimate of proportion correct (Swets, 1973), varying from 1.0 (perfect discrimination) to 0.50 (chance).

Table 1 shows the mean area scores for each participant group for each kind of Different trials, along with the standard deviation. The age groups did not differ from each other overall (mean = 0.74 for younger and 0.77 for older people), nor did age interact with any other factor. Musicians (mean = 0.80) outperformed nonmusicians (mean = 0.71), $F(1, 51) = 13.39$, $p < .001$), and overall people found the Rhythm and Contour items to be equally difficult (means = 0.83 and 0.82) whereas the Mode items were more difficult (mean = 0.61). Despite the hint from Experiment 1 that older musicians might classify the musical dimensions differently than the other groups, neither age nor experience interacted with item type, nor did all three factors interact (all *F*s near 1.0).

One thing of note in our results is the difficulty our participants had in distinguishing pairs identical except for Mode. Younger and older nonmusicians were essentially at chance in this comparison, but even musi-

TABLE 1

Mean Area Scores for Age x Experience Groups for Each Discrimination Type in Experiment 2

	Experience		
Discrimination Type	Nonmusician	Musician	Weighted Mean
Rhythm			
Younger	0.81 (0.13)	0.83 (0.11)	0.82
Older	0.79 (0.15)	0.90 (0.08)	0.85
Weighted mean	0.80	0.86	
Contour			
Younger	0.78 (0.11)	0.87 (0.12)	0.82
Older	0.75 (0.16)	0.89 (0.11)	0.82
Weighted mean	0.77	0.88	
Mode			
Younger	0.54 (0.15)	0.63 (0.16)	0.58
Older	0.56 (0.10)	0.71 (0.18)	0.64
Weighted mean	0.55	0.67	

Standard deviation in parentheses. Chance = 0.50. Means are weighted by the number of participants in each group.

cians did not perform well. To give a sense of the performance level on a more familiar scale, we also analyzed our data by dichotomizing the scale values into answers of "different" (scale values 1, 2, or 3) and "same" (scale values 4, 5, or 6). From this tabulation, we can report proportions of hits and false alarms. Hit rates were similar among all the groups (0.77 and 0.74 for younger and older nonmusicians, 0.80 and 0.79 for younger and older musicians). The false-alarm rate for younger nonmusicians was 0.67 to Mode trials, and older nonmusicians were the same at 0.68. The rate for younger musicians was 0.57, and for older musicians, although the most accurate of the groups, the rate was still 0.45.

Because of this surprising performance, we separated the older musicians into the most and least musically experienced. Our question was whether the very most experienced musicians in our sample would also have trouble with the mode distinction. Examination of the musical background questionnaires revealed that half the older musicians were highly trained professionals whereas the other half had more of an amateur status with respect to music performance. Formal analysis of this factor is precluded by the fact that only six people were in each group, and also by the fact that the average vocabulary score in the older professionals (33.7) was considerably higher than the score for the older amateurs (21.5 including one person with a score of 4; 25.0 excluding that person).

Nevertheless, we looked at hit and false-alarm rates for each subset of older musicians. The professional group indeed had higher hit rates than the amateurs (0.83 vs 0.73), and lower false-alarm rates (average false-alarm rate of 0.15 vs. 0.40). Nevertheless, it is interesting to note that both groups had similar ordering of false-alarm rates: for the professionals, the false-alarm rates for Mode, Rhythm, and Contour trials were 0.31, 0.13, and 0.02, respectively. The same means for the amateurs were 0.67, 0.30, and 0.23, comparable to the young nonmusicians.

DISCUSSION

With respect to the influence of age vs. experience on this task, the pattern is quite clear. Musicians exceeded nonmusicians in the ability to discriminate identical melodies from a melody differing in mode, rhythm, or contour. Although the advantage of experience seemed to be less prominent for rhythm (difference of 0.06 between the experience groups, vs. 0.11 and 0.12 for contour and mode, see Table 1), in fact experience and error type did not interact.

The story for age is also quite clear. Mean area scores did not differ for the age groups, with the numerical advantage actually in favor of the older listeners. Age did not interact with any of the other factors. We found no age by experience interaction suggesting smaller age differences among the more experienced listeners. In fact, whereas younger nonmusicians had essentially the same area scores as older nonmusicians (0.71 and 0.70), the younger musicians were actually numerically a bit worse than older musicians (0.78 and 0.83, respectively).

The ease of discriminating identical from changed pairs did differ reliably depending on the basis of the discrimination. Rhythm and contour changes were most easily detected, whereas mode was harder to detect. As noted earlier, this ordering of difficulty was the same for musicians and nonmusicians, as well as for older and younger people. Detection of mode was considerably harder than the other two discriminations. Nonmusicians could not really do this task at all, and the musicians were less than impressive in what we consider to be a very basic task, in that no interference or transformations other than a near-key transposition were imposed on the second melody. Even our most experienced musicians (musical experience gained by virtue of both age and career) made a considerable number of false alarms to different-mode pairs. We conclude that mode is not just a confusable dimension in memory but is an aspect difficult to assimilate even on-line in a perceptual task.

General Discussion

These two experiments speak to several interrelated points: the capabilities of nonmusicians in abstracting musical regularities, the differences between experience groups in that regard, the relationship between similarity measures and discrimination (and memory) tasks, the difficulty of the major/minor distinction, and the effect of age on processing of musical attributes.

As in our previous study (Halpern, 1984), in Experiment 1, nonmusicians seemed quite capable of abstracting the factorial nature of our stimulus set and rating similarity appropriately. By this, we mean that the clustering analysis returned exactly the factorial nature of the stimulus set, the fit of the factorial solution to the data was very high, and the solutions were similar to those of the musicians. The nonmusicians were able to do this despite the fact that the melodies were novel, the comparison required listeners to transpose, the melodies sounded similar to one another, and each melody pair was heard only once. This showing is superior to many of the cases cited by Smith (1997) in which nonmusicians showed weak categorical perception of intervals and chords, weak organization of the tonal hierarchy, and reacted primarily to aesthetic rather than formal aspects of music. Perhaps nonmusicians were able to perform so well here because the similarity ratings were an indirect way of reacting to the melody set. It is quite unlikely that nonmusicians could have articulated explicitly the organization of the whole stimulus set, but comparing only two melodies at a time may have reduced the cognitive load sufficiently to allow the nonmusicians to take advantage of their implicit musical knowledge. Also contributing to a low cognitive load was the fact that although the response period was limited to 6 s, this seemed more than sufficient for most listeners (and others so requesting received a little extra time). In addition although each pair was presented for judgment only once, each melody was heard four times in a session. This repeated exposure may have assisted the learning of the melodies.

A final reason nonmusicians may have been able to organize the stimulus set was that the rhythm and contour changes were both obvious ones. The two Rhythm categories were different in meter as well as rhythm, with Rhythm I being a duple and Rhythm II being a triple meter. Likewise, the contour was inverted in same-except-for-contour pairs because of their retrograde relation, which is the most dissimilar one contour can be from another (although it should be noted that people sometimes confuse a tune with its retrograde; Dowling ,1972). This may have exaggerated the dissimilarity of the rhythm and contour pairs and, by comparison, the similarity of the major/minor pairs.

The correspondence of the clustering solutions for the musicians and nonmusicians is notable (Figures 2 and 3), and the ratings of all the groups correlated fairly strongly. The exception to this is that the older musicians showed a tendency to group by contour first rather than rhythm first. But because the contour branches were not terribly distinctive in that solution, it is probably more accurate to say that they weighted more equally the distinctiveness of rhythm and contour than did the other three groups, who more clearly divided the set on rhythm first.

Turning now to the discrimination task, one thing that surprised us was that the task was not trivially easy for the musicians, many of whom had extensive training and experience performing on more than one instrument. Even for the rhythm and contour discrimination, area scores ranged between 0.83 and 0.90, certainly not ceiling performance.

The nonmusicians performed considerably more poorly on this task than did the musicians, although at an average area score of 0.71, they were still well above chance. Clearly this task is sensitive to musical experience across all the comparisons, as experience and pair type did not interact. So contrary to our hypothesis, experience did not especially aid what in our view was the most "musical" of the discriminations, that of mode. It may be the case that asking people to make the discrimination across the transposition put an extra premium on musical experience. It would be interesting to see if experience differences are diminished for rhythm and contour, but not for mode, if the compared tunes are at the same pitch level.

We should also note that the transposition selected here was to a musically close key; that is, the keys of F and C share many of the same pitches. This very closeness, however, has been shown to impede the discrimination of exact from near imitations. Bartlett and Dowling (1980) showed that discrimination of exact from near imitations improves across musically far key transpositions, such as a transposition between the keys of C and B major, which share few pitches. Perhaps transposition to a far key would have also improved the performance of nonmusicians, or enabled the musicians to use their background more effectively in the mode discrimination, leading to the larger experience difference we predicted for that task.

Generally speaking, the discrimination task results were congruent with the similarity rating data. Contour and rhythm pairs were rated both as being dissimilar and were discriminated at relatively high levels. It is indeterminate from these experiments whether perception of dissimilarity and the ability to tell the tunes apart are separate cognitive operations. It is possible that one is inclined to rate tunes as dissimilar because one can discriminate them, or perhaps one discriminates them well because they sound dissimilar.

The mode judgments are more informative in this regard. From the clustering analysis, we saw that major/minor pairs were rated as being highly similar. Similarity does not necessarily imply subjective identity, although it appears to in this case. The poor performance on the discrimination task shows us that for mode comparisons, the similarity ratings likely were due to the fact that the major/minor pairs sounded identical to the listeners. Thus, the memory confusions for mode found by Halpern (1984) and

Madsen and Staum (1983) could in fact have been caused by the perceptual confusion between major and minor shown in Experiment 2 here.

This perceptual confusion was shown even by the musicians, who had an area score of 0.67 for this discrimination. Although it is above chance, this is poor discrimination by any standard, and poorer still when the importance of mode in musical compositions is considered. The literature on affect and mode suggests that discrimination would have improved had we asked listeners to assign a "happy" or "sad" label to each tune before discrimination. Listeners could have adopted a strategy of silent labeling of this sort in the current experiment. Our evidence that they did not use it is of course indirect (the poor performance), but if our conjecture is correct, it is curious that the strategy was not spontaneously adopted. Perhaps the short unfilled retention interval prevented listeners from assigning an affect label to each melody separately. Alternatively or in addition, listeners may have poor metacognition in discrimination of mode, either by not realizing how poorly they are performing, or if so, not knowing how to improve (i.e., by assigning a label).

We should acknowledge at this point that the stimulus set used here, although it has several attractive features (the factorial structure, its use in a previous study), is of course a limited sample of all possible tunes and ways of defining similarity. We cannot be sure that a different set of tunes would return exactly the same results. However, in this regard, it is interesting to consider that two sets of tunes, constructed from different interval patterns, were used in the similarity ratings of the earlier study (Halpern, 1984, Experiment 1). The clustering solution showed that for neither musicians nor nonmusicians was interval pattern an important way of dividing up the melodies. All listeners first divided the group by rhythm and then by contour. This suggests that listeners' judgments were not all that influenced by the particular intervals of the melody but rather by the dimensions held in common across the different interval patterns.

Finally, we discuss the effect of age in this experiment. Experiment 1 showed that older and younger people heard similar structures in this unfamiliar music. In Experiment 2, we had expected that age effects would be more prominent for the global perceptual judgments of rhythm and contour, and smaller for the more musical judgment of mode, following from the results of our earlier study (Halpern et al., 1995). However, we found no overall effect of age and no interaction of age with type of discrimination. We cannot attribute this lack of effect to ceiling or floor performance (see Table 1), nor to overall insensitivity of our task, as experience effects were quite large. Instead, we think that once memory demands are minimized, as they were in both our experiments, we see that older adults are at least as good as younger adults in comparing music on some of its most fundamental dimensions.

Statistically, age and experience also failed to interact in any of our analyses. We did, however, have a hint that age and experience might interact if the very highest levels of musical background are required from subjects. In our case, the age x experience interaction tended toward a somewhat different form than is usually cited in the literature about aging. The usual form of an age x experience interaction is that younger experts perform the best, followed by older experts (e.g., Morrow, Leirer, Altieri, & Fitzsimmons, 1994). Age differences are larger among nonexperts, but again with superior performance by the young. Here, we had hints that our older experts were superior to younger experts. We are cautious about the conclusions from this analysis because of our small sample and the fact that our older professionals had very high vocabulary scores, which may reflect more schooling or a general intellectual superiority compared with the other groups. However, it is possible that the kind of musical experience accrued over 50 years of a professional career can compensate for normal age-related impairments in memory to such an extent that the experience overrides the biology. Experiments comparing older and younger performers with extensive professional experience would be useful in this regard.[1]

References

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

Bartlett, J. C., Halpern, A. R., & Dowling, W. J. (1995). Recognition of familiar and unfamiliar music in normal aging and Alzheimer's disease. *Memory and Cognition, 23,* 531–546.

Charness, N. (1989). Age and expertise: Responding to Talland's challenge. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 437–456), Cambridge, UK: Cambridge University Press.

Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics, 5,* 417–421.

Dowling, W. J. (1973). Rhythmic groupings and subjective chunks in memory for melodies. *Perception & Psychophysics, 14,* 37–40.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics, 50,* 305–313.

Dowling, W. J., Kwak, S., & Andrews, M. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57,* 136–149.

Gerardi, G. M., & Gerken, L. (1995). Development of affective responses to modality and melodic contour. *Music Perception, 12,* 279–290.

Halpern, A. R. (1984). Perception of structure in novel music. *Memory and Cognition, 12,* 163–170.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1995). Aging and experience in the recognition of musical transpositions. *Psychology and Aging, 10,* 325–342.

1. This research was supported by Grant RO1-AGO9965-01 from the National Institute on Aging. We are indebted to Paul Minda for assistance in preparation of materials, data analysis, and testing participants.

Halpern, A. R., Kwak, S. Y., Bartlett, J. C., & Dowling, W. J. (1996). The effects of aging and musical experience on the representation of tonal hierarchies. *Psychology and Aging, 11*, 235–246.

Jones, M. R., & Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Memory and Cognition, 19*, 8–20.

Krumhansl, C. L. (1991). Memory for musical surface. *Memory and Cognition, 19*, 401–411.

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 579–594.

Madsen, C. K., & Staum, M. J. (1983). Discrimination and interference in the recall of melodic stimuli. *Journal for Research in Music Education, 31*, 15–31.

Morrow, D. G., Leirer, V. O., Altieri, P. A., & Fitzsimmons, C. (1994). When expertise reduced age differences in performance. *Psychology and Aging, 9*, 134–148.

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

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*, 319–345.

Smith, J. D. (1997). The place of musical novices in music science. *Music Perception, 14*, 227–262.

Swets, J. A. (1973). The relative operating curve in psychology. *Science, 182*, 990–1000.

Welker, R. L. (1982). Abstraction of themes from melodic variations. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 435–447.