2004

Effects of training and melodic features on mode perception

A.M. Leaver

Andrea Halpern

Bucknell University, ahalpern@bucknell.edu

Follow this and additional works at: https://digitalcommons.bucknell.edu/fac_journ

Part of the Neuroscience and Neurobiology Commons

Recommended Citation

Effects of Training and Melodic Features on Mode Perception

AMBER M. LEAVER & ANDREA R. HALPERN
Bucknell University

The two modes most widely used in Western music today convey opposite moods—a distinction that nonmusicians and even young children are able to make. However, the current studies provide evidence that, despite a strong link between mode and affect, mode perception is problematic. Nonmusicians found mode discrimination to be harder than discrimination of other melodic features, and they were not able to accurately classify major and minor melodies with these labels. Although nonmusicians were able to classify major and minor melodies using affective labels, they performed at chance in mode discrimination. Training, in the form of short lessons given to nonmusicians and the natural musical experience of musicians, improved performance, but not to ceiling levels. Tunes with high note density were classified as major, and tunes with low note density as minor, even though these features were actually unrelated in the experimental material. Although these findings provide support for the importance of mode in the perception of emotion, they clearly indicate that these mode perceptions are inaccurate, even in trained individuals, without the assistance of affective labeling.

Received July 15, 2003; accepted February 2, 2004

Music consists of complex combinations of spectral and temporal features. Patterns of these features, and especially deviations from patterns already established within a piece of music or culture, convey important information that can elicit a vast range of images and emotions. Although some amount of expertise is needed to create such musical motifs, it seems that one does not need any special faculties to experience and appreciate them. Presumably, the accomplished concert pianist and the “tone-deaf” individual both know when to expect the man-eating shark in Jaws—they knew even when they heard that ominous theme for the very first time.

The notes, or pitches, of music are the primary conveyors of emotion or expressiveness. Rhythm does have the ability to convey some information...
regarding levels of intended expressiveness (Kendall & Carterette, 1990) and some types of emotion (happiness and fear, Juslin & Madison, 1999). However, these effects are most meaningful when they interact with pitch information (Schellenberg, Krysciak, & Campbell, 2000; Kendall & Carterette, 1990; Juslin & Madison, 1999).

More specifically, the mode that pitches convey is an effective communicator of mood. The mode of a piece of music refers to the set of pitches, or scale, from which the piece is derived. Although a large number of modes are possible, most of Western music is written in major or minor mode, which draw from the major scale and minor scale, respectively. Music written in major mode is typically associated with feelings of happiness and lightheartedness, whereas minor music is often considered to be more somber and sad.

This association between mode and the conveyance of emotion is well documented in the music perception literature (Crowder, 1985; Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Hevner, 1935; Kastner & Crowder, 1991; Whissell & Whissell, 2000). One of the first to document this phenomenon, Hevner (1935) found that participants used more positive adjectives to describe major melodies and selected more negative words to describe minor melodies. Other researchers have shown that even young children reliably assign positive affective labels to major melodies and negative affective labels to minor melodies (in this case a happy or sad face; Dalla Bella et al., 2001; Gerardi & Gerken, 1995; Kastner & Crowder, 1991). In fact, Kastner and Crowder (1991) showed this occurring as early as 3 years of age. Whissell and Whissell (2000) investigated mode as a composer’s tool, finding that Beatles tunes written in minor keys contained significantly more negative lyrics than those written in major keys. Clearly mode plays an important role in conveying emotion.

Despite the strong association between mode and affect, research on mode perception has uncovered perceptual difficulties in both nonmusicians and the musically inclined. Halpern (1984), for example, explored mode perception as compared with rhythm and contour perception. She constructed two highly structured melody sets, each consisting of variations on one core melody. These melodies differed from their core melody in one of the three features: rhythm, contour, or mode (for a more explicit description of the construction of the melody set, please refer to the Methods section of Experiment 1).

In the first task, musicians and nonmusicians rated every possible pair combination both within and between each melody set for similarity. Elements in each pair group were presented in different keys so that participants could not rely on changes in absolute pitch, but were instead forced to use the actual musical feature. Both musicians and nonmusicians
grouped melodies more frequently by similarities in rhythm, followed by contour, then mode. In a second task, participants identified melodies using labels (letters). After the learning phase, both musicians and nonmusicians confused melodies that differed only in mode most often, followed by rhythm, then contour. Taken together, these results showed that melodies differing only in mode were not only perceived as similar (similarity rating task), they were often perceived as being the same melody (identification task).

In a later experiment, Halpern, Bartlett, and Dowling (1998) examined these same issues using a similar paradigm. Replication of the similarity rating task using only one melody set found contour more distinguishing than rhythm, but mode was least often used to group the melodies. Halpern et al. (1998) also conducted a discrimination task using all pairwise combinations differing by only one feature—rhythm, contour, or mode. Although musicians performed better than nonmusicians overall, there was a consistent pattern to performance across all groups, irrespective of training. Rhythm and contour discrimination was equally good, but mode discrimination was poor for both musicians and nonmusicians. Participants did not consistently distinguish major and minor melodies.

Other researchers have also noted inaccuracies in mode judgments. For example, Pechmann (1998) has shown that discrimination of the mode of single chords seems to be inconsistent, even in highly trained individuals. In the relevant part of this experiment, a chord was presented, followed by a series of distracter tones, and finally a target chord. Participants, students of a music conservatory, were asked whether the target chord was the same as the initial chord. When the target and reference chord differed only by mode, performance was above chance but poor (error rates just below 40%).

In a study by Dewitt and Crowder (1986), participants unselected for musical background discriminated between a tune and either an exact transposition of that tune (“same”), a transposition differing by mode only (“different”), or a transposition that retained mode but did not retain contour (“different”). Even when interstimulus interval was short, opposite mode pairs were labeled as the same (an incorrect response, $M = .90$) just as frequently as exact transposition pairs (a correct response, $M = .89$). Clearly, these participants heard pairs of melodies differing only in mode as being the same melody.

The following experiments attempt to elucidate the nature of mode perception by exploring the relationship between mode and affect and the relative deficits in mode classification and discrimination in both musicians and nonmusicians. Experiment 1 borrows the paradigm used by Halpern and her colleagues (Halpern, 1984; Halpern et al., 1998) in order to generalize their results to a different melody set. This experiment also adds
affective and musical classification of single melodies to the mode discrimination task. Experiment 2 concentrates exclusively on mode discrimination and classification in nonmusicians, examining the effects of task order. Experiment 3 explores the effects of experimental and natural musical training on mode perception. The final experiment explores the role that other nonmode melodic features play in inhibiting accurate mode perception.

Experiment 1

We constructed a melody set similar to that used by Halpern and her colleagues (Halpern, 1984; Halpern et al., 1998) in order to compare mode discrimination with contour and rhythm discrimination. In addition to the discrimination task, participants also performed a classification task, where they were asked to classify members of a new 40-item melody set using the labels “happy” and “sad” or “major” and “minor.” We hypothesized that those participants who excel on one task should also do so on the other, if both tasks tap a common ability to process mode.

Nonmusician participants were selected because they showed the greatest deficits in prior studies of mode perception, and lower baseline scores would allow any successful manipulation to be more visible. Nonmusicians may or may not have been familiar with the concept of mode prior to this experiment, but it is probably safe to assume that all participants were familiar with the concepts of happiness and sadness. For that reason, we hypothesized that classification performance would be facilitated by affective labels. Also, as Halpern et al. speculated, we expected the availability of affective labels to improve discrimination performance. Although affect was not explicitly linked to mode discrimination during the experiment, we did expect that participants who performed affective classification first should perform at higher levels on the subsequent discrimination task (compared with the reverse order). We thought that classification would both call attention to and allow practice of labeling the exemplars. A successful discrimination strategy could then consist of labeling (major/minor or happy/sad) each member of the pair and using the labeling results to drive the discrimination decision. Finally, we predicted that minor melodies would be more difficult to classify than major melodies, for several reasons. First, minor melodies are less common in popular Western music. Second, the tonal hierarchies (systems of note relatedness) established by major and minor modes are not only different in structure (Krumhansl & Kessler, 1982), but also differ in strength such that major mode seems to establish the key or tonal hierarchy more strongly than minor mode (Krumhansl, Bharucha, & Kessler, 1982;
Krumhansl & Kessler, 1982; Harris, 1985). Last, there are multiple types of minor scales used in Western music (natural, harmonic, and melodic) compared with one major mode, which could also make minor classifications more complicated.

METHOD

Participants

Forty-three (33 female and 10 male, mean age = 18.5 years, SD = 0.55 years) Bucknell University students were recruited to participate in return for partial fulfillment of a course requirement. All were nonmusicians as assessed by a musical background survey. Musical experience in voice or on an instrument (as defined by serious practice, private lessons, and/or participation in an ensemble) was no greater than 3 years and was on average much less than that maximum (M = 0.29 years, SD = 0.32 years). Musical experience occurred a mean of 6.41 years (SD = 2.54 years) before the experiment.

Materials

The construction of the discrimination task melody set replicated the factorial nature of the Halpern (1984) melody sets. The bank of melodies was based on one core melody. This original series of pitches was first reversed, creating two orders with completely opposite contours (retrogrades). These two pitch patterns were then fitted into two different rhythmic patterns both in triple meter creating four different melodies. These melodies were then changed to the minor mode of the same key (major to minor). Thus, a set of melodies were created in such a way that minimal pairs could be made that differed in only one melodic feature—either contour, rhythm, or mode. However, to ensure that participants would make mode discrimination judgments based on the melodic feature and not differences in absolute pitch, pair elements were presented in different tonic pitches (C and F). These 40 melody pairs were presented with an interstimulus interval of 4 s and an intertrial interval of 6 s. In all experiments, melodies were synthesized so that quarter note length was 0.5 s with minimal spacing between notes. All melodies in the discrimination task were 5 s in duration. Participants heard each melody pair only once, for a total of 40 trials per task. The stimuli are presented in the key of C in Figure 1.

These and all subsequent stimuli were created within certain parameters to convey a strong tonal sense. First, all melodies started with the tonic pitch and ended with either the leading tone or the second scale degree moving to the tonic. Second, we attempted to maximize the number of notes belonging to the tonic triad (e.g., C, E, and G in the key of C.

Fig. 1. Factorial melody set used in Experiment 1.
major), especially on stressed beats. Finally, the perceived musicality of all stimuli was assessed by four musically trained observers (at least two in all subsequent experiments).

Stimuli used in the Classification Task consisted of 40 obscure melodies taken from volumes of old folk tunes that fulfilled the musicality parameters just described. This set contained tunes of various lengths (\( M = 8.5 \text{ s}, \text{SD} = 1.67 \text{ s} \)) and tonic pitch (key). The melodies contained an average of 22.1 notes and were varied in tempo, but were all within a moderate range (quarter note \( \approx 0.5-0.6 \text{ s} \)). Twenty of these melodies were in minor mode, the other half were in major mode. Participants heard each melody only once, yielding 40 trials for this task. The melodies were presented in two different random orders, separated by a 4-s intertrial interval. Tunes in both tasks were played in a piano timbre through MIDI output speakers of a Yamaha PSR 500 keyboard mediated by Cakewalk Professional 5.0 software.

**Procedure**

Each participant performed the discrimination task and classification task, but in different orders and using one of two different classification systems. Thus, there were four groups. Approximately half the listeners were administered the discrimination task first followed by the classification task; the other half did these tasks in the reverse order. Half of each of these groups used affective labels (“happy” and “sad”) in the classification task, whereas the other half used musical labels (“major” and “minor”). Participants used a 6-point confidence rating scale in both tasks, with 1 meaning “sure different” and 6 meaning “sure same” in the discrimination task.

Before the tasks, the experimenter provided four example melodies, two major and two minor, that were not part of the test melody set. If participants indicated that they could hear no difference between major and minor, the example melodies were played again. After this, if participants were still unable to hear a difference, they were asked to do the best they could to classify the melodies (only a few participants reported not being able to hear a difference between example melodies; these were all participating in major/minor, not affective, classification and their scores were not noticeably different from other participants in the major/minor condition).

**RESULTS**

**Discrimination Task**

In the main analysis of the discrimination task, confidence ratings were converted into area under the receiver operating curve (AUROC) scores (Swets, 1973). Perfect discrimination was equivalent to a score of 1.0, and chance performance was equivalent to 0.5.

In order to examine the effects of task order and classification type on discrimination performance, a 3 (feature type) \( \times \) 2 (task order) \( \times \) 2 (classification type) analysis of variance (ANOVA) was applied to the data. There was, as expected, a significant main effect of feature type, \( F(2, 84) = 38.60, p < .001 \). Post hoc analysis (Tukey’s honestly significant difference) showed that although rhythm and contour discrimination were not different (\( M = .72 \) and .71, respectively, SD = .17 for both), mode discrimination was significantly lower (\( M = .54, \text{SD} = .13 \)). The 95% confidence intervals of this mode discrimination score ranged from .50 to .58 and thus can be considered chance performance. There were no significant effects for either order or classification type.
There was a significant interaction between feature type and order, $F(1, 42) = 7.08, p = .01$. Rhythm discrimination was improved when performed after the classification task ($M = .76, SD = .14$ when last; $M = .68, SD = .18$ when first), whereas contour discrimination declined ($M = .67, SD = .16$ when last, $M = .73, SD = .18$ when first). However, mode discrimination performance was still poor regardless of whether it was done before ($M = .53, SD = .13$) or after ($M = .55, SD = .11$) classification.

Despite this discrepancy in performance across feature type, there was no difference in the confidence of the participants’ responses. The rating scale was converted from 6 to 3 points, with “3” being maximum and “1” being minimum confidence. This allowed for analysis of the overall confidence of the response, regardless of its accuracy. Overall, participants showed a low level of confidence, with an average of 1.28 ($SD = 0.48$) for exact transpositions (“same” pairs), 1.31 ($SD = 0.59$) for rhythm, 1.40 ($SD = 0.55$) for contour, and 1.11 ($SD = 0.51$) for mode pairs. There was no significant difference in confidence for discrimination of any pair type $F(3, 180) = 2.32, p = .08$.

**Classification Task**

Classification confidence ratings were dichotomized, with ratings of 1, 2, and 3 indicating a response of “major” or “happy” and 4, 5, and 6 indicating “minor” or “sad,” depending on the condition. In this way, we could measure performance in terms of percent correct for major and minor tunes. A 2 (task order) $\times$ 2 (classification type) $\times$ 2 (mode) ANOVA revealed a significant effect of classification type, $F(2, 42) = 57, p < .0001$. Participants using the affective rating scale classified melodies more accurately ($M = .84, SD = .07$) than did participants using the major/minor classification labels ($M = .55, SD = .17$). The limits of the 95% confidence intervals for mean classification scores using major/minor labels included .50 (upper bound = .61, lower bound = .49). No other main effects or interactions were significant.

**Discrimination and Classification**

There were no significant correlations between performance on either condition of the classification task and performance on any portion of the discrimination task.

**DISCUSSION**

As in Halpern et al. (1998), mode discrimination was more difficult than contour or rhythm discrimination. Also as predicted, affective labels facilitated mode classification. However, although we predicted that par-
participants who performed affective classification first might have used those familiar happy/sad labels to differentiate melodies in the discrimination task, we found performance in these tasks to be unrelated. The affective classification task apparently did not encourage listeners to employ emotional labels to assist in the discrimination task.

Also, performance on mode classification using either label type did not correlate with mode discrimination performance. Even those who did well in classification still performed poorly (or inconsistently) in discrimination. Considering that classification performance using affective labels was quite high and that mode discrimination was at chance, a correlation, if it existed, may have been obscured due to restriction in the range of scores.

Interestingly, the major tunes in this set of stimuli were just as difficult to classify as the tunes written in minor mode. This was true regardless of the labels participants used. If melodies written in minor mode are more harmonically ambiguous or less common, it did not affect classification performance in this task.

**Experiment 2**

Experiment 1 attempted to introduce implicit training in the form of affective labels; in Experiment 2, we attempted to implement implicit training in a different way by highlighting mode differences in the discrimination task. Instead of pairs differing by any one of three features, here the number of melodic features that could differ was decreased to one, mode. We hypothesized that removing the other two features (contour and rhythm) might facilitate performance both within that task and in the classification task as well. Because mode classification scores using affective labels were already relatively high, participants in this experiment classified using only musical labels. In this way, if these new manipulations had an effect, it would be more apparent. Ultimately, we hoped that implicit training in the form of increased exposure to same-except-for-mode melody pairs and task order would boost performance on this new mode discrimination task. This training should also benefit classification when it follows discrimination.

**METHOD**

**Participants**

Twenty (12 female and 8 male, mean age = 18.75 years) Bucknell University students participated in this experiment in return for partial fulfillment of course credit. Again, all participants were nonmusicians, with no more than 2 years of experience (M = 0.21 years, SD = 0.25 years) occurring an average of 6.55 years (SD = 2.14 years) before the experiment.
Materials

The stimuli used in the classification task of Experiment 2 were identical to the stimuli used in Experiment 1. However, the discrimination task was different from that of Experiment 1. The new discrimination stimuli consisted of 32 obscure folk tunes. By using extant folk melodies, we were better able to ensure musical validity and to allow for the possibility of generalization of Experiment 1’s results across more melodies. Each of the 32 melodies was rewritten in the mode opposite to the original melody. This way, each tune had an equally musical opposite-mode partner. Each melody was also transposed to the keys of C and F. These melodies varied in contour and meter but were all within a similar moderate tempo range. Tunes contained an average of 21.3 notes and were on average 8.04 s (SD = 1.72 s) in length. Pairs of these melodies were constructed such that pair elements differed either by key only (same except for transposition) or by key and mode (different). As in Experiment 1, participants heard each pair once, yielding a total of 32 trials for this task.

Participants heard one of four sets of stimuli. These were counterbalanced such that each core melody was heard in every possible mode pair-type combination (major-major, minor-minor, major-minor, and minor-major) at some point throughout the four lists. The order of key pair-types (C, F; F, C) was counterbalanced as well, but not for each core melody. Instead, each mode combination was presented in every possible key combination within each set, but not across every core melody.

Procedure

Each participant performed both the mode discrimination and classification tasks, but approximately half of these (n = 9) did discrimination first, while the rest (n = 11) did classification first. The classification task remained exactly as it was in Experiment 1, except that participants classified melodies using only the musical labels of “major” and “minor.” During both tasks, participants were asked to use a 6-point confidence scale.

RESULTS

Mode Discrimination Task

AUROC scores were used as the dependent measure in this analysis, as in Experiment 1. Performance was at chance regardless of whether the task was completed first or second (M = .50, SD = .10 for both), t(18) = .164, p = .87.

Classification Task

Two-way ANOVA (Task Order × Mode) revealed no main effect of task order, F(1, 36) = 2.66, p = .11, despite the fact that those participants who performed the classification task first scored somewhat lower (M = .59, SD = .16) than did those who did this task second (M = .69, SD = .19). The main effects of mode and the interaction were also not significant.

Mode Discrimination and Classification

There were small positive correlations between mode discrimination and mode classification performances, but none were significant.
**Experiment 1 vs. Experiment 2**

It seems clear that task order did not have an effect within the second experiment, but did mode discrimination in the presence or absence of discrimination of other feature types affect performance on either task? We ran an ANOVA on mode discrimination AUROC scores (three factors: task order, classification type, and discrimination type) as well as percent correct classification (two factors: task order and discrimination type) across both experiments.

Against our prediction, there was no main effect of discrimination type. Mode discrimination in Experiment 2 was no higher ($M = .50, SD = .10$) than mode discrimination in Experiment 1 ($M = .54, SD = .13$), $F(2, 60) = .96, p = .33$. There were no significant effects or interactions across any other factors for mode discrimination.

Performance in the classification task using musical labels (major/minor) was slightly improved in Experiment 2 ($M = .64, SD = .17$) over Experiment 1 ($M = .55, SD = .17$). This difference approached but did not reach significance, $F(1, 39) = 3.58, p = .07$.

**DISCUSSION**

Our predictions were not supported by the results of Experiment 2, except that major and minor melodies were equally easy to classify. Despite the alterations made to the mode discrimination task, performance was still quite poor. Highlighting mode by requiring only discrimination of pairs differing by mode did not facilitate that discrimination or classification of the mode of single melodies. This was true even though these stimuli were taken from volumes of folk tunes and not composed within the experimental constraints used to create the discrimination stimuli of Experiment 1. The correlations between discrimination and classification performance, although positive and possibly overshadowed by overall low discrimination performance, were not significant. It seems the implicit training we attempted to build into this experiment was not successful in improving performance in mode discrimination as compared with Experiment 1, or in increasing performance on the second task performed in Experiment 2.

Experiments 1 and 2 both indicate that the perception of mode, a powerful communicator of emotion in music, is inconsistent in nonmusicians. Major and minor melodies were only classified well when the labels of “happy” and “sad” were used. Performance using the musical labels of “major” and “minor” was much lower, as was discrimination performance in all conditions.

Clearly these participants hear the distinction between modes, as they are able to appropriately label these melodies using affective vocabulary.
Low classification performance using musical labels could be explained by a lack of familiarity; perhaps nonmusicians simply do not understand what “major” and “minor” mean, despite being exposed to some training and examples, and therefore they cannot use the labels accurately. However, mode discrimination performance is troublesome. Participants were certainly familiar with the concepts of “same” and “different.” If participants can hear the distinction, why was discrimination performance so low?

Smith, Kemler Nelson, Grohskopf, and Appleton (1994) faced a similar dilemma. In their study, participants identified and discriminated between three different intervals (a major third, minor third, and perfect fourth) and slightly mistuned versions of those intervals. Two of these intervals, the major and minor third, form the first interval of the major and minor chord, respectively. All three intervals also happen to correspond to the openings of three well-known folk tunes. Nonmusician participants who were informed of this folk tune connection were better able to identify the three intervals than participants who were not given this information. However, this availability of the folk tune labels in Smith et al. (1994) was not helpful during the discrimination task. Availability of the folk tune labels clearly affected classification performance in Smith et al., as was the case in the affective classification task in Experiment 1 of this article. Perhaps explicit training in the differences between major and minor melodies would increase classification performance using “major” and “minor” labels by making specific instances more available in memory. Experiment 3 explored the issue of training.

**Experiment 3**

In this experiment, nonmusician participants performed classification and discrimination tasks identical to those of Experiment 2, after receiving different types of training: (1) affective, (2) music theory, or (3) combination (affective and music theory). The affective training group was taught the link between mode and affect, associating major and minor melodies with light-hearted and somber moods, respectively. Members of the music theory training group were instructed in mode from the standpoint of music theory, involving a discussion of major and minor scales. A third group (combination) received both types of instruction. After instructional training, all groups entered into a brief feedback session during which participants classified three major and three minor melodies. Musicians were included to provide a comparison with sustained training. Finally, all participants performed both the classification and discrimination tasks of Experiment 2.
Overall, we expected that attaching more meaning of any type to mode would facilitate performance on both tasks. We also expected training groups to perform at different levels. The combination group should have an advantage over other groups, just by virtue of having more content. Affective training should be more effective than music theory training for the same reason we expected that musical labels would be more difficult than affective labels to apply in the classification task. Although participants may not have heard of scales or mode before the experiment, they would be likely to be familiar with these two basic emotional states.

Directly applying the results of Smith et al. (1994), training linking affective and musical labels should aid classification, but not discrimination. However, Smith et al.’s tasks (musical interval judgments) were different from those presented in this article. It may be easier to apply an emotion to a melody than to link an interval to a familiar melody. Not only do intervals and mode judgments differ in complexity, Smith et al.’s task involved three musical intervals, whereas this experiment deals with only two different modes. For these reasons, the results of Smith et al. may not successfully predict the effect of training on mode discrimination; training might improve mode discrimination. However, in light of successful application of affective labels in Experiment 1 and because mode discrimination seems to be more difficult than classification, telling participants that major melodies are “happy” should have a greater effect on classification than discrimination performance.

**METHOD**

**Participants**

Fifty-nine Bucknell University students were recruited to participate in this experiment in exchange for partial fulfillment of course credit. Forty-five were nonmusicians (26 female and 19 male, mean age = 18.9 years) with 0.47 years of musical experience on average (SD = 0.46 years) occurring an average of 5.90 years (SD = 2.85 years) before the experiment. Fourteen were musicians with an average of 9.82 years of music lessons (SD = 2.33 years). Only two of these musicians were not musically active at the time of the experiment.

**Materials**

Both the stimuli and the method of presentation were identical to Experiment 2.

**Procedure**

All nonmusicians received training before the experiment. The affective group was instructed in the link between mode and affect. They were told that major melodies are typically associated with light-hearted or happy moods, and minor melodies with somber or sad moods. The music theory group was instructed in the link between scales and melodies. These participants were told that major melodies take their notes from major scales, whereas minor melodies take their notes from the minor scale. The third group, combination, was given both types of training. All three groups heard the same four exam-
ple melodies during training; music theory training juxtaposed examples of major and 
minor scales in addition to these example melodies. The same feedback session occurred 
after training in all groups, where six melodies were played (three major and three minor) 
and subjects were asked to respond whether each was major or minor. Musicians were not 
given extra training, but given instructions and example melodies identical to those given 
in Experiment 2. After training, participants completed classification and discrimination 
tasks, which were counterbalanced identically to Experiment 2.

RESULTS

Mode Discrimination Task

As in Experiments 1 and 2, AUROC scores were the dependent measure used in analysis. Untrained nonmusicians from Experiment 2 were included in this analysis for the sake of comparison. A one-way ANOVA comparing performance across the five participant groups (three training groups, untrained nonmusicians from Experiment 2, and musicians) showed a significant difference among these means, $F(4, 71) = 6.50, p < .001$ (means are summarized in the bottom half of Table 1). Post hoc analysis (Tukey’s honestly significant difference) revealed significant differences between the untrained nonmusicians and both the musicians ($p < .01$) and the combination group ($p < .05$). No other comparisons between groups were significant.

Contrast analyses also showed positive effects of long- and short-term training. A contrast of untrained to all trained nonmusicians was significant, $F(1,71) = 15.05, p < .001$. A contrast of musicians to all nonmusicians was also significant, $F(1,71) = 9.55, p = .003$.

Classification Task

A two-way ANOVA (Training Group × Mode) showed a significant main effect of group, $F(4, 71) = 7.50, p < .0001$ (means summarized in

<table>
<thead>
<tr>
<th>Training</th>
<th>Classification Labels Used</th>
<th>Discrimination</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No training, nonmusician</td>
<td>Happy/sad</td>
<td>.54 (.16)</td>
<td>.84 (.07)</td>
</tr>
<tr>
<td></td>
<td>Major/minor</td>
<td>.53 (.10)</td>
<td>.55 (.17)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No training, nonmusicians</td>
<td>Major/minor</td>
<td>.50 (.09)</td>
<td>.64 (.17)</td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Music theory, nonmusicians</td>
<td>Major/minor</td>
<td>.62 (.18)</td>
<td>.67 (.18)</td>
</tr>
<tr>
<td>Affective, nonmusicians</td>
<td>Major/minor</td>
<td>.64 (.13)</td>
<td>.82 (.10)</td>
</tr>
<tr>
<td>Combination, nonmusicians</td>
<td>Major/minor</td>
<td>.67 (.16)</td>
<td>.79 (.15)</td>
</tr>
<tr>
<td>No training, musicians</td>
<td>Major/minor</td>
<td>.74 (.15)</td>
<td>.88 (.16)</td>
</tr>
</tbody>
</table>

Note—Values in parentheses represent the standard deviation. Discrimination scores for Experiment 1 reflect mode discrimination performance only.
Table 1), but not mode. The interaction was also not significant. Post hoc analysis revealed significant differences between several groups. Nonmusicians who received no training and those receiving music theory training performed equally poorly on this task. Both these groups performed significantly lower than the affective training group (p < .01 and p < .05, respectively), and musicians (p < .01 for both). Untrained nonmusicians also had lower scores than the combination group (p < .05); training groups did not differ from one another.

Similar to classification task analysis, a contrast comparing untrained nonmusicians against the three training groups was significant, F(1,96) = 10.11, p = .002, as was a contrast of musicians against all nonmusicians, F(1,96) = 10.24, p = .002.

**Mode Discrimination and Classification**

Performance on the two tasks was positively correlated for all groups. These correlations were significant for nonmusicians receiving affective training, r(13) = .54, p = .04, and for those that received music theory training, r(13) = .66, p = .01. Correlations approached, but did not reach, significance for nonmusicians receiving both types of training, r(12) = .48, p = .08, and for musicians, r(10) = .51, p = .09.

**DISCUSSION**

Instruction, either in the form of several years of experience or short experimental sessions, facilitated both the classification and discrimination of mode. As the contrast analyses showed, the long-term experience musicians brought to the tasks was more beneficial than any training provided by the experiment. These analyses also revealed that experimental training was better than no training. Again, major and minor melodies were equally easy to classify.

Despite these general findings, our hypotheses regarding the specific effects of training on each task were only partially supported. We hypothesized that the combination of affective and music theory training would confer the greatest advantage, followed by affective training and music theory instruction, in that order. This hypothesis was not supported by classification scores, but was by the pattern of discrimination performance. Mean discrimination performance was indeed highest in the combination group, followed by the affective, music theory, and untrained group. However, not all differences between these groups were significant. Only the combination group was significantly better than the untrained group.

Classification scores only partly followed the predicted trend. In classification, the affective group scored highest, followed by the combination,
music theory, and finally untrained groups. However, not all differences between groups were significant. Only affective instruction significantly increased classification performance above untrained performance.

Although nonmusicians performed at a higher level on mode discrimination when trained than when not, scores were still low when compared to classification scores. In fact, scores in both the classification and discrimination tasks were relatively low overall. Although musicians scored better than nonmusicians, their mean discrimination score (.74) was still not close to ceiling. Classification scores were similarly lower than expected, especially for musicians (.88). We expected at least musicians to approach ceiling performance on these simple tasks. Individuals with only a modest amount of formal musical background are exposed early and often to the idea of mode, in the form of exposure to music written in the two different modes, explicit training, and drills involving major and minor scales and other exercises. Even most introductory music methods books include tutorials on mode. Individuals with a moderate musical background should have had sufficient exposure to mode to excel at these tasks.

Because performance did not reach our expectations, we began to wonder what could be interfering with mode judgments. We looked for stimulus features that systematically affected mode judgments. More specifically, we checked for positive correlations between correct classification and the strength or prevalence of different musical features, including the number of notes per second (note density), average pitch height, and the relative number of times different scale degrees were sounded (measured by percentage of total notes). We did find positive correlations between ratings of “major” and high note density and high percentage of the third degree of the scale (which is the most consistent note change between major and minor). This was true despite the fact that major and minor melodies did not differ with respect to those features during instruction, training, or in either task. These correlations seemed to be strongest in musicians and in nonmusicians receiving combination instruction. Based on these preliminary analyses, we devised a new experiment that focused specifically on the effects of these two melodic features on mode classification: note density and the prevalence of the third scale degree.

Experiment 4

In order to attempt to explain relatively poor performance in these mode perception experiments, Experiment 4 focused on two musical features and their role in mode classification. We manipulated major and minor tunes so that they were either high or low in note density (number
of notes per second) and either high or low in the percentage of notes of
the tune that were the third of the scale (the note that defines major and
minor mode).

We used note density as our measure for melody speed because, while
both tempo and note density are related to speed, tempo is relatively arbi-
trary. Note density, on the other hand, is absolute. For example, the typi-
cal tempo of an American march is 120 bpm, but one could change that
tempo to 60 by changing the meter. Regardless of the meter and tempo
chosen, note density remains the same, which makes it the more unam-
biguous choice.

Based on the post hoc analyses of Experiment 3, those tunes high in
note density should be more likely to be classified as major than tunes low
in note density. Empirical studies have already linked higher tempos,
another musical feature related to speed, with positive affect and slower
tempos with negative affect (Gagnon & Peretz, 2003; Hevner, 1936,
1937; Rigg, 1940; Scherer & Oshinsky, 1977). Thus, performance should
be higher when the affective information conveyed by mode and note den-
sity agree (high-density major tunes and low-density minor tunes) and
lower when affective messages do not agree (high-density minor tunes and
low-density major tunes).

The second musical feature, the third degree of the scale, plays an
important role in the mode of a melody. It is largely the third scale degree
(e.g., E vs. E♭ in C major and C minor) that conveys mode information.
Therefore, melodies that have a relatively high percentage of thirds (e.g.,
E's or E♭'s) should yield more accurate performance than those with a low
percentage.

Musical training and exposure to mode should also affect how sensitive
participants are to these nonmode melodic features. We have shown pre-
viously that participants with little or no formal musical training are not
as sensitive to mode as are participants with formal training. Therefore,
untrained nonmusicians should not show sensitivity to these other melod-
ic features. However, nonmusician participants who receive experimental
training should have a better, yet still incomplete, conception of mode and
should be more likely to use a strategy, however inappropriate, to make
mode judgments. These participants, especially because they would be
made aware of the link between mode and affect, should be more suscep-
tible to the effects of note density, as our post hoc analysis showed them
to be in Experiment 3. Our prediction about musicians was somewhat less
straightforward. On the one hand, our post hoc analysis indicated that
musicians' scores were also influenced by these nonmode melodic fea-
tures. On the other hand, musicians should be much less likely to be
"fooled" by conflicting affective messages conveyed by note density, if
they are influenced by this feature at all, because of their independent
understanding of mode. Because the materials in Experiment 3 were not constructed to explore note density as a possible specious indicator of mode, the materials in Experiment 4 were carefully designed to help us tease these factors apart.

**METHOD**

**Participants**

Forty-three (31 female and 12 male, mean age = 19.62 years, SD = 1.39 years) Bucknell University students were recruited to participate in this experiment in return for partial fulfillment of course credit. Twenty-five were nonmusicians with no more than 3 years experience (M = 1.97 years, SD = 1.51 years) occurring 6.39 years (SD = 2.69 years) before the experiment on average. Eighteen were musicians with an average of 10.06 years (SD = 1.50 years) of experience.

**Materials**

The melody set used in this classification task was derived from that used in the discrimination tasks of Experiments 2 and 3. Two melodies were discarded and four were altered so that tunes, which ranged from 1.67 to 3.19 notes per second, could be split at the group’s median (2.46) into “high” (M = 2.87, SD = 0.23) and “low” (M = 2.13, SD = 0.24) note density (notes per second) groups. Tunes were also split at the median score (.22) for percentage of the third scale degree. These scores ranged from .08 to .38, and were separated such that half had a “high” percentage of the third degree of the scale (M = .29, SD = .04) and the other half was “low” in that percentage (M = .16, SD = .04). As in the discrimination tasks of the prior experiments, sets of 30 melodies were counterbalanced so that each tune was heard in each key (C or F) and in each mode (major and minor) across each group. Each melody was heard only once, totaling 30 trials for this task.

**Procedure**

Musicians and one group of nonmusicians received no training, other than the instruction and examples given to participants in Experiment 2. The other group of nonmusicians was given combination (affective and music theory) training used in Experiment 3. The musician and combination groups were chosen because their patterns of classification showed the strongest correlations with the two features in question. Untrained nonmusicians served as a baseline comparison group. All participants completed a classification task using the stimuli described earlier.

**RESULTS**

As in Experiment 3, training had a positive effect on mode. A 3 (training) × 2 (mode) × 2 (note density) × 2 (percentage of third scale degree) ANOVA showed a significant main effect of training, F(2, 23) = 26.38, p < .001. Post hoc analysis (Tukey’s honestly significant difference) indicated significant differences between all groups. Musicians scored higher than trained (p = .02) and untrained (p < .001) nonmusicians. Trained nonmusicians (combination training) scored higher than untrained nonmusicians (p < .001). No other main effects were significant. All relevant means are displayed in Table 2.
There was a significant interaction between note density and mode, $F(1, 23) = 6.43, p = .01$. Pairwise comparisons at $p < .05$ showed that while note density affected classification of minor melodies, $F(1, 156) = 6.70, p = .01$, it did not have an effect on major melodies, $F(1, 156) = 1.00, p = .32$. However, classification scores for major melodies did follow the predicted trend; performance for low-density major melodies ($M = .70, SD = .22$) was lower than that for high-density major melodies ($M = .74, SD = .19$). Classification of low-density minor melodies ($M = .75, SD = .17$) was significantly more accurate than classification of high-density minor melodies ($M = .66, SD = .18$).

There was also a significant Training $\times$ Note Density $\times$ Mode interaction, $F(2, 23) = 3.34, p = .04$. Pairwise comparisons at $p < .05$ indicated that this interaction was not significant for untrained nonmusicians or musicians, although scores followed the predicted trend for musicians. However, performance of trained nonmusicians was affected significantly by note density. Trained nonmusicians classified major melodies with high note density more accurately than they classified major melodies with low note density, $F(1, 156) = 4.43, p = .04$, and they classified minor melodies with low note density more accurately than they classified minor melodies with high note density, $F(1, 156) = 8.50, p = .004$.

No other interactions were significant. The percentage of the third scale degree did not have an effect on classification performance, nor did performance differ for major or minor tunes overall.

**DISCUSSION**

Experiment 4 replicated the results of Experiment 3 in regards to training. Musicians received the highest scores, followed by nonmusicians receiving combination instruction, and finally nonmusicians who did not receive experimental training. Major and minor melodies were equally difficult to classify, which is also consistent with the results of earlier experiments.

---

**TABLE 2**

**Mean Classification Scores for Different Feature Types, Experiment 4**

<table>
<thead>
<tr>
<th>Note Density</th>
<th>Untrained Nonmusicians</th>
<th>Trained Nonmusicians</th>
<th>Musicians</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Major: .60 (.09), Minor: .57 (.13)</td>
<td>Major: .82 (.24), Minor: .62 (.20)</td>
<td>Major: .80 (.15), Minor: .78 (.12)</td>
</tr>
<tr>
<td>Low</td>
<td>Major: .61 (.22), Minor: .59 (.14)</td>
<td>Major: .69 (.25), Minor: .80 (.14)</td>
<td>Major: .82 (.22), Minor: .86 (.11)</td>
</tr>
<tr>
<td>% Third scale degree</td>
<td>High: .56 (.16), Minor: .60 (.12)</td>
<td>Major: .76 (.22), Minor: .72 (.17)</td>
<td>Major: .83 (.13), Minor: .77 (.12)</td>
</tr>
<tr>
<td>Low</td>
<td>Major: .65 (.17), Minor: .57 (.15)</td>
<td>Major: .75 (.28), Minor: .70 (.22)</td>
<td>Major: .79 (.16), Minor: .87 (.11)</td>
</tr>
</tbody>
</table>

*NOTE—Values in parentheses represent the standard deviation. Trained nonmusicians received combination instruction.*
Classification performance was not affected by the prevalence of the third degree of the scale, which is the most consistent pitch difference between major and minor mode. Despite the fact that attending to thirds should have helped mode judgments, participants were not sensitive to this manipulation. It is possible that the range of thirds was not wide enough to induce a difference. Regardless, this feature cannot explain lower than expected performance on mode discrimination in these experiments.

In contrast, although attending to note density should not have helped mode judgments, some participants nevertheless used note density to classify mode. This was true despite the fact that each melody was heard in both modes, and therefore major and minor melodies did not differ in respect to note density overall. Trained nonmusicians were especially influenced by note density. For this group, when mode and note density conveyed the same emotion (positive: high-note-density major tunes, negative: low-note-density minor tunes), accuracy of mode classification increased. When the affective messages conflicted (low-note-density major tunes, high-note-density minor tunes), performance suffered. That these participants were influenced by note density lends further support for the instability of the mode concept in these individuals.

Untrained nonmusicians were not sensitive to this manipulation, which is not surprising given that they were not particularly sensitive to mode in this task either. Musicians were also not affected by the manipulation of note density, which could be indicative of their more complete understanding of mode. However, considering that analysis of all groups showed an overall effect of note density on minor classification, it is also possible that the note density range was too limited to elicit the desired effect. The range of affect that this particular selection of note densities conveyed was not tested before the experiment, so it is possible that we have captured a medium to high range of note densities (which is the perception of the authors). If that is the case, this could explain the lack of effect on major melodies. To be certain of this possibility, we would either need to increase the range of note densities or test the range of affect conveyed by these stimuli.

**General Discussion**

The experiments presented in this article indicate that the perception of major and minor mode is inaccurate in individuals with and without musical experience. It is curious that what is such an important communicator of the emotional and expressive content of Western music is processed selectively. Sensitivity to this feature seems to lag behind sensitivity to other melodic features (rhythm and contour, Experiment 1). Even
performance of musicians in Experiments 3 and 4 was lower than expected. We expected that musicians would approach ceiling levels on what we thought were trivial tasks.

However, despite disappointing performance on direct tests of mode perception, various indirect tests did show sensitivity to mode. Highlighting the link between mode and affect and introducing training did increase the accuracy of mode judgments. In addition, successful interference of a nonmode melodic feature that also conveys affect with mode judgments in Experiment 4 showed that participants do have a concept of mode as a communicator of emotion. Although mode may not be as attentionally salient as one would think, people are not wholly insensitive to the differences between major and minor.

MODE AS A MELODIC FEATURE

The first finding presented in this article was that mode discrimination is poorer than discrimination of rhythm and contour, reinforcing the same finding reported previously by Halpern (1984) and Halpern et al. (1998). What differs between the discrimination of rhythm and contour and the discrimination of mode that would make the latter so much more difficult?

One possibility is that the magnitude of change within rhythm and contour melody pairs, at least in these experiments, is greater than between elements of a mode pair. Mode differences involve a local change of only a few pitches, while rhythm and contour differences change the entire melody. However, those few notes that differ between major and minor mode create a global change in emotional content, as indicated by classification performance using affective labels. Regardless of whether the differences in these three features are equal in magnitude, low mode discrimination performance is still troubling.

Discriminating rhythm and contour involves comparing and matching patterns, whereas mode discrimination requires assigning a category label to melodies. Intuitively, it may seem easier to discriminate mode because participants need only to assign one of two labels to each melody. However, it is possible that abstracting the mode from a melody is more complicated than simply comparing patterns. Certainly mode discrimination is different in this way from rhythm and contour discrimination, which may explain the marked differences in performance in discrimination of these features.

THE LINK BETWEEN MODE AND AFFECT

The results reported here provide evidence for a strong link between mode and affect. In Experiment 1, untrained nonmusicians who were not able to reliably discriminate between pairs of melodies differing by mode
were able to classify major and minor melodies using the labels of “happy” and “sad.” Classification performance using affective labels in this experiment was dramatically higher than classification performance of untrained nonmusicians in Experiment 1 using the music labels of “major” and “minor.” In Experiment 3, training that explained this link between mode and affect (affective and combination) boosted performance on both classification and discrimination tasks. Clearly, even those who have very little musical training assign positive affect to major melodies and negative affect to minor melodies, just as prior studies have shown.

Further evidence for the link between mode and affect comes from the effects of note density on mode judgments. Note density, another melodic feature that carries emotional content, interfered with the perception of mode in both Experiment 3 (post hoc analyses) and Experiment 4 (experimental manipulations). When the messages from these two communicators of affect conflicted, performance suffered, especially for trained nonmusicians (combination). A conflicting message from this nonmode melodic feature was never enough to bring performance below chance (e.g., participants never consistently responded “minor” to a major melody with low note density), but the message was strong enough to bring performance significantly below cases where messages did not conflict. That this effect occurred for both major and minor modes adds strength to the effect. Although the finding that note density was able to interfere with mode perception indicates an incomplete understanding of mode, it also adds more support to the idea that mode conveys emotional content in music.

**Mode Classification vs. Mode Discrimination**

Intuition suggests that classification and discrimination use similar processes. Both tasks require holding two input items in one's working memory: a target stimulus and a reference stimulus. In classification, the reference stimulus might be an internal, prototypical representation instead of an experimenter-supplied stimulus (as in a discrimination task). However, in both cases, the number of items is the same.

A series of empirical studies also support the idea that classification and discrimination are similar. Nosofsky (1983) showed that discrimination of auditory signals differing in intensity was not different from classification of those signals. Wilson, Wales, and Pattison (1997) reported a significant positive correlation between the classification and discrimination of tonal and atonal melodies in 7- and 9-year-old children. This evidence suggests that there are no major differences in task demands between classification and discrimination.

However, the experiments presented in this article show major performance discrepancies between the classification and discrimination of mode.
Mode discrimination scores were consistently much lower than those for mode classification in all experiments. Even when discrimination scores were at their highest in Experiment 3 (.74 for musicians), they still lagged behind classification scores (.88 for the same group). Discrepancies in performance between these two tasks suggest that processing demands differ between classification and discrimination, making one more difficult than the other.

If classification and discrimination are similar, then performance on these two tasks should be correlated. Experiment 3 did indeed show positive correlations between classification and discrimination. However, these correlations were significant for only some groups. In addition, at classification and discrimination performance was not significantly correlated in Experiments 1 or 2. Moreover, if classification and discrimination tested the same process, task order should have affected performance. To the extent that there are similarities between these task types, participants should have applied whatever knowledge they used to classify melodies to enhance discrimination performance. However, discrimination scores were equally low regardless of whether classification was completed first. Again, classification and discrimination seem to be tapping different processes.

A number of things could make mode discrimination different from mode classification. While in classification the reference stimulus is internally represented and well learned, in discrimination the reference stimulus is being updated with each new stimulus pair. In this way, discrimination may tax working memory more than classification. In addition, the discrimination task requires comparison of transpositions, further complicating the task. These things might complicate discrimination enough to make this task more difficult for participants than mode classification, thus explaining the discrepancies in performance on these tasks.

The structure of experimental training may have conferred more advantage to classification performance than to discrimination performance. Training emphasized the classification of major and minor melodies without directly comparing melodies differing only by mode. Moreover, as one of our reviewers pointed out, it did not make affect explicitly relevant to the discrimination task. Experimental training could be altered such that discrimination judgments were highlighted, and participants could be asked to discriminate on the basis of the emotional content of the melodies. Although we were hoping that they would do this implicitly, these changes would likely improve mode discrimination performance.

THE EFFECTS OF EXPERIMENTAL TRAINING AND MUSICAL BACKGROUND

The results of the final experiments indicate that training of any type increases performance on tasks of mode perception. In Experiment 3, non-
musicians performed significantly more poorly than musicians and experimentally trained nonmusicians on both the classification and discrimination task. In Experiment 4, nonmusicians lagged behind musicians and experimentally trained nonmusicians yet again. Whether the instruction is short-term or long-term, participants clearly benefited from training.

Turning to long-term training, several analyses showed that musicians outperformed all nonmusicians. The musical background developed by years of long-term training gave these participants an advantage when making mode judgments. It is also possible that the superior performance of musicians is the result of a natural proclivity that led them to study music in the first place. However, in light of the effectiveness of experimental training in the later experiments reported in this article, it seems more likely that it is training and experience—not innate musical ability—that is most advantageous for tasks of mode judgments.

In some cases, experimental instruction was effective enough to bring performance levels up to the performance level of musicians. For the classification task, musicians' performance was no different from the performances of the affective and combination nonmusician training groups in Experiment 3. Moreover, for the discrimination task of the same experiment, musician, affective, and music theory training groups all performed equally well. Depending on the type of training and task performed, the 6 or 7 min of training given in this experiment were just as effective as the several years of musical background the musicians brought to the table.

Musicians recruited to participate in these experiments were all young college students, most of whom participated in musical ensembles in middle school and high school. Many were still actively participating in musical activities at the time of the experiment, but many were doing so at a reduced level. However, although they certainly had more experience than nonmusician participants, they were not professional musicians. It may be more accurate to refer to them as "moderately trained" or "amateur" musicians. If similar experiments were conducted on professional musicians, or even music majors, performance would likely reach the ceiling levels we originally expected.

Halpern et al. (1998) actually tested a group of older musicians. This group had a mean area score of .71 (SD = .18) on their discrimination task using factorial melody sets of the type used in Experiment 1 of this article, which was similar to the performance level of Musicians of Experiment 3 in this article (M = .74, SD = .15). When Halpern and her colleagues separated professional from amateur musicians in this group of older individuals, hit rates for the group of professionals (.83) were higher than the hit rates for amateurs (.73), but the more experienced group was still not completely accurate in mode discrimination. So although the musicians of Experiments 3 and 4 of this article may not be "professional" musicians, the results of Halpern et al. suggest that their performance
is no more inaccurate than groups that have even more extensive musical backgrounds.

**IMPLICATIONS: INCOMPLETE REPRESENTATIONS IN NONEXPERT KNOWLEDGE**

Despite the fact that musician participants may not have performed differently than trained nonmusicians in some conditions, musicians do seem to have a different, and perhaps more complete, understanding of mode. In Experiment 4 musicians were less susceptible to the interference of note density than the group of nonmusicians that also showed this interference effect (combination training). That musicians were not "fooled" by conflicting affective content as supplied by the note density feature indicates that they have a better grasp on mode. In fact, this may have contributed to their higher overall performance on this task.

Research in expert knowledge has also shown that although in some cases performance of experts and nonexperts may not differ quantitatively, their performances do differ qualitatively. For example, in a study by Ericsson and Harris, participants who were not familiar with chess were tested for memory for patterns of pieces on a chessboard (as cited in Ericsson & Charness, 1994). With training, nonexpert participants were able to achieve performance as high as that of chess masters. However, their strategies differed. Whereas nonexperts focused on patterns in the periphery that were not relevant to chess, chess masters used patterns in the center of the board relevant to move selection. This study shows a fundamental difference in strategy and cognitive processing between experts and nonexperts—a result that has been found in other domains as well, including typing, memory, sports, and others (Ericsson & Charness, 1994). So, although nonmusicians receiving experimental training in Experiments 3 and 4 may have performed as well as musicians did, nonmusician performance was qualitatively different from the performance of musicians. Experiment 4 indicates that nonmusicians rely more heavily on affective information to make mode judgments.

**MAJOR AND MINOR MELODIES ARE EQUALLY EASY TO PROCESS**

Interestingly, our hypothesis that minor melodies would be more difficult to classify than major melodies was consistently refuted throughout all experiments. Despite the fact that minor melodies are less common in popular music than major melodies, and despite the fact that minor mode tends to be more harmonically ambiguous than major mode, participants found either mode equally difficult to classify, regardless of condition. It is possible that the minor melodies used in these experiments were not as harmonically ambiguous as typical minor melodies. The stimuli used in these experiments were chosen explicitly because they were relatively
stereotyped, so the minor stimuli used here may not be representative of the typical melody written in minor mode. Without implementing experimental manipulations specific to the differences between major and minor, it is difficult to draw definitive conclusions. However, the evidence presented in these experiments suggests that major and minor are equally easy to process for both musician and nonmusician participants.

The only difference in performance between major and minor modes occurred in Experiment 4, where an analysis of all groups showed a significant interaction between note density and mode for minor, but not for major, melodies. However, as mentioned previously, it is unclear whether this reflects a real difference between major and minor or is caused by the restricted range of note density of the stimuli. After this work was completed, Gagnon and Peretz (2003) published a study that found an interaction of tempo and mode in the perception of emotion similar to that found in Experiment 4. Those authors found the interaction to be significant for both major and minor melodies, and they also noted that tempo seemed to have a stronger effect on emotion than mode. In light of their findings, we suspect that including melodies with lower note density would indeed expand the effect of note density to both major and minor melodies.

**PRACTICAL APPLICATIONS FOR MUSIC EDUCATION**

Based on the results of this series of experiments, the prospect of effectively training individuals to successfully classify and discriminate mode seems a difficult task. Individuals with moderate amounts of musical background did not perform as well as expected on mode judgments. Even individuals with substantial musical backgrounds are not entirely accurate in mode perception, as evidenced by Halpern et al. (1998) and Pechmann (1998). However, that experimental training introduced in Experiments 3 and 4 was somewhat successful is heartening. Although experimental training did not increase performance even to that of musicians, its duration was only several minutes. Increasing the duration of the training session, including a longer feedback session and allowing for more interaction with the material being taught, would very likely increase performance even more.

It seems that making participants aware of the link between mode and affect is more successful than giving more musical explanations of major and minor mode. On the classification task, nonmusician participants receiving affective instruction outperformed all other groups of nonmusicians, and their scores were no different from the scores of musicians. For discrimination, however, this type of training was no more successful than any other types. It is likely, though, that because this task was more difficult, any effects of training may have been obscured. Explaining the link
between mode and affect seems to be vital for the understanding of major and minor mode.

Because the link between mode and affect is so important, it would be advantageous to control for other musical features that convey emotion in Western music when training participants in the differences between major and minor mode. The results of Experiment 4 indicate a case where an affect-bearing musical feature, note density, was able to “fool” some participants into giving melodies the mode label corresponding to the affective message communicated by the nonmode feature (e.g., labeling a minor melody with high note density as “major”). If individuals are made aware of other melodic features and the emotional content they carry, they may be less likely to be “fooled” by conflicting affective messages when making mode judgments. In fact, training individuals using example melodies with conflicting affective messages could be quite effective. So although performance on the tasks of mode perception presented in this article was consistently inaccurate for both musicians and nonmusicians, the implications of these results for the future direction of mode education are promising.1

References

1. We would like to thank Christopher Plaue for his help in computing the AUROC scores.

This paper is based on a master’s thesis written by the first author under direction of the second author. Portions were presented at the meeting of the Eastern Psychological Association, Baltimore MD, March 2003.