IS THE BANDWIDTH FOR TIMBRE INVARIANCE ONLY ONE OCTAVE?

—

KENNETH M. STEELE AND AMBER K. WILLIAMS
Appalachian State University

TIMBRE INVARIANCE REFERS to the ability to determine whether two notes at different pitches were played or sung by the same instrument or voice. Handel and Erickson (2001) reported that nonmusician listeners heard pairs of notes as coming uniformly from different instruments when the pitches were separated by an octave or more; they concluded that the bandwidth of timbre invariance was only one octave. Here we replicate that study with methodological refinements and include musicians as well as nonmusicians. We presented listeners with pairs of notes from two instruments (horn and bassoon) spanning a 2.5 octave range, and listeners judged whether two notes were the same pitch and produced by the same instrument. Nonmusicians replicated Handel and Erickson's result of a decline in timbre invariance beyond 1.0 octave, whereas musicians' performance declined less—to about 80% correct at 2.5 octave. Pitch judgments did not vary across the range and were more accurate for musicians than for nonmusicians. The difference between musicians and nonmusicians in timbre judgments suggests caution in stating a range for the operation of timbre invariance.

Is the Bandwidth for Timbre Invariance Only One Octave?

TIMBRE is defined as the quality by which two notes may be judged to be dissimilar when pitch and loudness have been equated (ANSI, 1973). This definition tells little about the nature of timbre itself, only that it is not subsumed by pitch or loudness. Hadja, Kendall, Carterette, and Harshberger (1997) reviewed timbre research methodology and concluded that there is still a great deal of mystery about timbre as a musical variable. Psychologically, it is unclear whether timbre should be measured on a categorical or continuous scale. The physical correlates of timbre are incompletely defined, although it is clear that they consist of more than steady-state harmonics alone (Iverson & Krumhansel, 1993; Risset & Wessel, 1999).

People treat timbre as a categorical variable in daily use, identifying timbre with the source instrument. Three different notes on the piano keyboard have individual timbres, since each played note produces a unique combination of transients and harmonics. However, people will treat them as having the common quality of being piano notes, distinct from, for example, clarinet notes. Handel and Erickson (2001) used the phrase *timbre invariance* to describe the quality that is assumed to be preserved across the notes of an instrument. Handel and Erickson asked the experimental question of the range, or bandwidth, of timbre invariance by investigating a listener's ability to determine whether two notes at different pitches came from the same instrument or singer.

The first experiment of Handel and Erickson (2001) investigated how well listeners identify whether two notes at different pitches were played by the same instrument or not. Their participants were musically naive undergraduates. The note range was from G3 to C6 and the notes came from six wind instruments, recorded on the McGill University Master Samples Compact Disks (Opolko & Wapnick, 1987, 1989). Each note consisted of the first 1.5 to 2.5 s of the original recordings, with a smoothed offset. Notes were normalized to the same amplitude. On every trial, two notes were presented in an ABABA sequence. The lowest pitch was presented first, and there was a 0.5 s silence between each note. The participants responded on a four-point scale ranging from "(1) very certain identical instrument" to "(4) very certain different instruments."

Handel and Erickson reported that average judgments could be collapsed into two categories, note pairs with less than one octave separation and note pairs with more than one octave separation. Note pairs separated by more than one octave were judged as coming from different instruments uniformly. Handel and Erickson concluded that one octave of separation of notes constitutes a border, beyond which timbre invariance breaks

down. Handel and Erickson reported supporting evidence in a second experiment that compared the ability of experienced and inexperienced listeners to identify classically trained singers at different pitches.

Handel and Erickson's conclusion is important because it suggests that the everyday assumption that all notes from an instrument have some common identifiable quality is wrong. Their conclusion is startling because it suggests that the range of invariance is very narrow, only one octave in width. However, there are concerns about their procedure which should be investigated before their conclusion is accepted.

One issue is the lack of counterbalancing of experience with instruments and notes. Participants had more experience with trials where the separation was less than one octave ($n = 45$ trials) or equal to one octave ($n = 41$ trials) than with note pairs spanning more than one octave ($n = 27$ trials). Thus, lack of experience is correlated with the judgments that participants found most difficult. Participants had more experience with different-instrument pairs ($n = 78$ trials) than with identical-instrument pairs ($n = 35$ trials), which may have created a response bias toward judging instruments as different when uncertain. Finally, different octave separations were confounded with instrument differences. For example, when the first note of a pair was C4, the number and type of instruments varied for the following comparisons: C4/G4 (6 instruments), C4/C5 (4 instruments), and C4/C6 (3 instruments). The C4/G4 comparison may have contained easier discriminations that resulted in greater accuracy.

A second issue concerns the dependent measure. The participants responded on a four-point scale ranging from "(1) very certain identical instrument" to "(4) very certain different instruments." One of the concerns above was that the unbalanced number of identical- and different-instrument trials may have created a response bias toward different-instrument responses when the participant was uncertain. Since there is no measure of the accuracy of judgment, the changes across octave separations could be interpreted as due to changes in response bias as the discrimination task became more difficult.

We chose to replicate the first experiment from Handel and Erickson with modifications of their procedure to meet the above concerns. In addition, we chose to compare a musically skilled group against a less-skilled group since musical expertise may modulate the effect. Beal (1985) and Pitt (1994) found that musicians could treat pitch and timbre as separable dimensions in discrimination tasks while nonmusicians had great difficulty identifying pitch when timbre varied also. Pitt concluded that nonmusicians were biased to attend more to timbre and hypothesized that pitch was controlled by the mix of harmonics. For nonmusicians, the pitch of a note would be increased by the occurrence of higher harmonics in the note. The same note would be of two different pitches if the harmonics were sufficiently different.

In this experiment, musician was defined as a person, majoring in music, who had completed at least one aural skills training class. Aural skills students are trained to identify pitches and to be able to reproduce them vocally or on another instrument. We hypothesized that successful aural skills training required that the student was able to separate pitch from timbre. Nonmusician was defined as a person who had not had an aural skills training class and was not majoring in music.

Participants were presented with two notes, presented in an ABABA sequence, that could differ in pitch, instrument source, or both conditions. Participants were asked to identify whether the notes had the same pitch or not, whether the notes were produced by the same instrument or not, and to provide confidence ratings of their judgments. Handel and Erickson (2001) would predict a strong decline in accuracy of identification of instrument identity as notes become separated by one octave or more. The work of Beal (1985) and Pitt (1994) suggests that these effects should be reduced in people with music training.

Method

Participants

Seventy-six (76) undergraduate students participated in the study: 38 students were introductory psychology students and 38 students were music majors, having completed at least one semester of aural skills training. Introductory psychology students participated to satisfy a course requirement. Music majors were recruited through music organizations and were paid $5 in compensation.

Apparatus

Note pairs were recorded on compact disc (CD), played on a Sony CDP-XE500 CD Player, amplified through a Onkyo TX08211 stereo receiver, and transmitted to Advent AW720 wireless headphones.

Procedure

Students were recruited to participate in groups of four per session. After informed consent was obtained, an

instruction sheet and a set of calibrated headphones were presented to each participant. Participants were told that they would listen to 56 note pairs, presented in ABABA sequence, and that they would be asked four questions about each pair of notes. The first question asked whether the notes were identical or not. The second question asked the participant to rate the degree of certainty of this judgment on a 1 (*very uncertain*) to 5 (*very certain*) scale. The third question asked whether the instruments were identical or not, and the fourth question asked the participant to rate the degree of certainty of this judgment on a 1 (*very uncertain*) to 5 (*very certain*) scale.

Participants listened to 4 note pairs to hear the range of differences that would be encountered (same note/same instrument, different note/same instrument, same note/different instrument, different note/different instrument). The participants were informed that notes separated by an octave (e.g., C3 and C4) were considered different notes. This last instruction was directed toward music majors because of two possible interpretations of the same-note question.

Notes from a French horn and a bassoon were obtained from a musical-instrument sample database maintained by the Electronic Music Studios at the University of Iowa. These instruments were chosen because of the range of overlap of their chromatic scales and because they were from different instrument families. The loss of the ability to discriminate between instruments from different families was seen as a fair test of the timbre invariance concept. The instruments were recorded in the anechoic chamber in the Wendell Johnson Speech and Hearing Center at the University of Iowa using a Neumann KM 84 cardioid condenser microphone, a Mackie 1402-VLZ mixer, and a Panasonic SV-3800 DAT recorder. The recordings were digitally transferred and edited into sound files consisting of chromatic scales. All samples are in mono, 16-bit, 44.1 kHz, AIFF format. Each note is approximately 2 seconds long and is immediately preceded and followed by ambient silence. The samples, and additional information, may be obtained from http://theremin.music.uiowa.edu.

Six notes in half-octave steps (C2, F#2, C3, F#3, C4, F#4) were selected from the two instruments. Tone onsets were preserved and the twelve notes were normalized to the same approximate volume. A 100 ms fade-out was applied to the end of each note. The ABABA sequences were assembled with a 0.5 s period of silence at the beginning, between each note, and at the end of the sequence. Note pairs were constructed in a low-pitch–high-pitch sequence following Handel and Erickson (2001). There are 78 possible note combinations, but pilot tests showed that participants had difficulty maintaining attention over a session that presented all combinations. Therefore all F#-C and F#-F# comparisons were removed, leaving the C-C and C-F# comparisons. The C2-F#4 comparison set was repeated twice to balance the number of comparisons at the larger octave separations. These changes produced a set of 56 comparisons, consisting of separations of 0 octave ($n = 12$), 0.5 octave ($n = 12$), 1.0 octave ($n = 8$), 1.5 octaves ($n = 8$), 2.0 octaves ($n = 8$), and 2.5 octaves ($n = 8$). The order of instruments was counterbalanced across the sequence.

Results

The results in Figure 1 show the accuracy for musicians and nonmusicians of classification of instrument identity as a function of the octave separation among notes. Chance performance (50% correct) is indicated by a dashed line. Overall, musicians ($M = 89.8\%$ correct) were more accurate than nonmusicians ($M = 61.8\%$ correct), $F(1,74) = 142.5$, $p < .001$. Both musicians and nonmusicians showed a decline in accuracy when the octave separation was increased. The performance of nonmusicians declined to chance levels when notes were separated by more than one octave, an effect consistent with Handel and Erickson (2001). Musicians showed a decline of about 10% in instrument identification accuracy across the range of octave separations, but overall accuracy was always above 80%. These results suggest that musicians show a wider range of timbre invariance.

The results in Table 1 show the pattern of errors in the instrument identification task as a function of octave
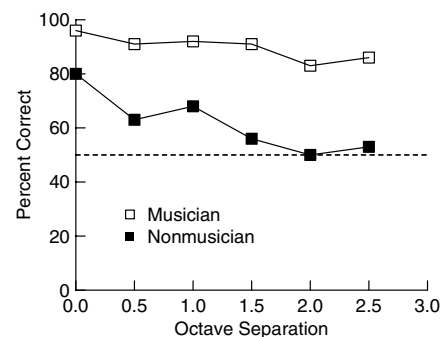

FIG. 1. Instrument identification accuracy. Each point indicates the mean percent correct for identification of whether the same instrument produced both notes at a particular octave separation. Musicians and nonmusicians are presented separately. The dashed line indicates a chance level (50%) of accurate identification. Standard error bars are obscured by the data points.

TABLE 1. Error patterns in instrument identification task

| | | Incorrect Choice | |
Major	Octave	"Same"	"Different"
Music	0	16	1
	0.5	33	9
	1.0	12	13
	1.5	6	20
	2.0	12	39
	2.5	8	36
Not Music	0	88	3
	0.5	138	31
	1.0	61	36
	1.5	37	96
	2.0	35	117
	2.5	18	124
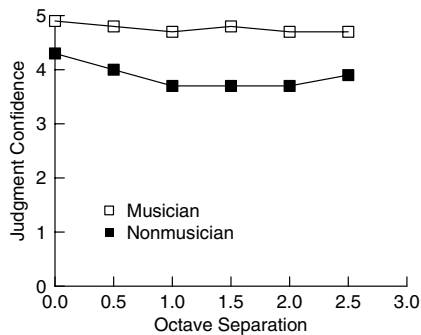
FIG. 2. Instrument identification confidence. Each point indicates the mean level of confidence of accuracy of judgment whether the same instrument produced both notes at a particular octave separation of the note pairs. Standard error bars are obscured by the data points.

separation and type of major. The case of "Same" indicates a judgment by the participant that the notes were played on the same instrument when, in fact, the notes were played on different instruments. Both music majors and nonmajors showed the same pattern of changes as the octave separation between notes was increased. Participants misjudged instruments as being identical when notes of identical pitch were played (0 octave separation). There was a systematic shift to misjudging instruments as being different when the octave separation between notes was increased. This shift toward hearing identical instruments as different would be consistent with a loss of timbre invariance.

The results in Figure 2 show confidence ratings by musicians and nonmusicians of their performance in the instrument identification task. Overall, musicians ($M = 4.8$, $SD = 0.3$) were more confident of their judgments than nonmusicians ($M = 3.9$, $SD = 0.7$) and this difference was statistically significant, $F(1,74) = 47.5$,

$p < .001$. It was notable that nonmusicians maintained high confidence in their judgments despite their chance levels of performance at the greater octave separations. This result suggests that self-judgments cannot be a safe substitute for objective measures of discrimination accuracy.

Handel and Erickson (2001) reported that their results could be collapsed into single categories of correct discrimination below one octave of separation and lack of discrimination above one octave of separation. Their results suggested a sharp categorical loss of timbre invariance. The pattern of results in Figure 1 suggests a graded loss of timbre invariance. Furr and Rosenthal (2003) described an extension of contrast analysis that permits one to compare the relative fit of multiple patterns to observed data. Briefly, the analysis works by assessing the degree to which each individual participant manifests an expected pattern of results, and then the difference in fit for different patterns is computed at the level of the individual. Finally, a significance test and an effect size are computed for the group of participants. Table 2 shows the outcome of fits of both a step function and a linear function to the instrument identification data. The results in Table 2 indicate that instrument identification was consistent with both a linear function and a step function, with the effect being stronger in nonmusicians. The difference score tests the difference in fit, with a positive score indicating a better fit by the linear function and a negative score indicating a better fit by the step function. There was no significant difference in fit between the two functions, suggesting that neither pattern uniquely described the results.

The results in Figure 3 show the accuracy of classification of note identity as a function of octave separation of notes for musicians and nonmusicians. Chance performance (50% correct) is indicated by a dashed line. Overall, musicians ($M = 93.3\%$ correct) were more accurate than nonmusicians ($M = 78\%$ correct), $F(1,74) = 42.8$, $p < .001$. Both musicians and nonmusicians showed little decline in accuracy when notes were more separated in pitch. The lack of effect of octave separation on note identification is in contrast to the separation effect in the instrument identification task. Also worthy of comment is the decline in accuracy when the notes were separated by 0.0, 1.0, and 2.0 octaves. The number of errors when the judgment involved equal chroma ($n = 404$) was twice the amount those involving only tone height differences ($n = 196$). When the two notes were identical in chroma and in the same octave, 96.3% (157/163) of mistakes were produced by notes from different instruments.

TABLE 2. Pattern analysis of instrument identification accuracy.

Major	Linear Pattern			Step Pattern			Difference Score		
	t	p	r_c	t	p	r_c	t	p	r_c
Music	3.75	.001	.52	3.54	.001	.50	1.58	.12	.25
Nonmusic	9.05	<.001	.83	8.52	<.001	.81	−.86	.40	−.14

Note. A higher *t*-value indicates a better fit between the pattern and obtained results. Effect size is indicated by r_c ($r_{contrast}$). Unstandardized contrast weights were 5, 3, 1, −1, −3, −5 for the linear function and 1, 1, 1, −1, −1, −1 for the step function. Contrast weights are standardized to compare across patterns. Negative difference scores indicate a better fit by the step function.
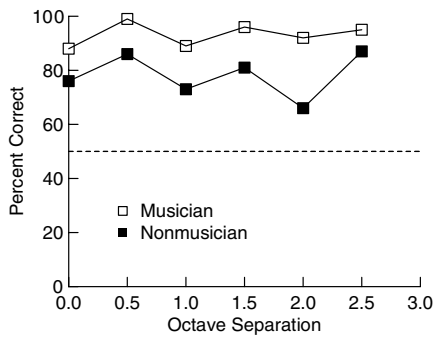

FIG. 3. Note identification accuracy. Each point indicates the mean percent correct for identification of whether the notes were identical pitches. Musicians and nonmusicians are presented separately. The dashed line indicates a chance level (50%) of accurate identification. Standard error bars are obscured by the data points.
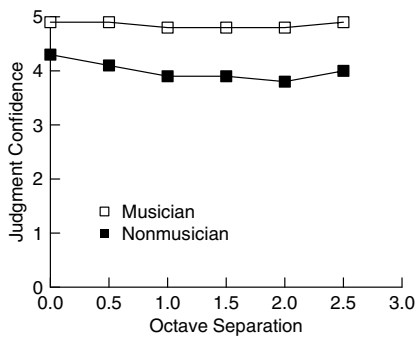

FIG. 4. Note identification confidence. Mean level of confidence of accuracy of judgment whether the notes were the same pitch. Each point indicates the overall mean confidence score whether the notes were the same pitch as a function of octave separation. Standard error bars are obscured by the data points.

The results in Figure 4 show confidence ratings by musicians and nonmusicians of their performance in the note identification task. Overall, musicians ($M = 4.8$, $SD = 0.3$) were more confident of their judgments than nonmusicians ($M = 4.0$, $SD = 0.6$) and this difference was statistically significant, $F(1,74) = 52.8$, $p < .001$. Confidence in their judgments were unaffected by differences in octave separation for both musicians and nonmusicians.

Discussion

The basic results of Handel and Erickson (2001) were replicated. Nonmusicians had great difficulty identifying whether two notes were produced by the same instrument, or not, when the notes were separated by one or more octaves. The pattern of errors for all participants showed a shift toward misidentifying notes as coming from different instruments when the octave separation was increased. However, musicians showed only about a 10% decline in accuracy and were able to maintain above an 80% accuracy rate at all separations. Thus musicians were able to maintain timbre invariance across a broader range than nonmusicians.

Why was there no sharp decrease in performance for the musically trained listener? The obvious answer is better listening skills. Our results may be partially explained by the conclusions of Pitt (1994) and Beal (1985) that musicians are able to treat pitch and timbre differences separately. However, musicians and nonmusicians showed similar patterns in their common mistakes. Both groups misjudged different instruments as being identical when notes of identical pitch were played, and both groups misjudged identical pitches as being different when they were played on different instruments. Finally, both groups showed an increase in misjudgments when pitches were the same chroma but in different octaves.

One notable aspect of the performance of the nonmusicians on the instrument identification task was that confidence in judgments remained high even though accuracy had declined to chance levels. This distinction between accuracy and confidence in one's accuracy is termed the *feeling of knowing* effect in the cognitive literature (Hart, 1965; Nelson, Gerler, & Narens, 1984). The literature suggests that two components of our procedure would bias participants toward feelings of high confidence despite low accuracy. Confidence in one's accuracy is increased by familiarity with the procedure and by lack of feedback regarding accuracy (Perfect & Hollins, 1999). Participants may have confused their

confidence of their knowledge of the steps of doing the task with confidence in how well they are doing the task. Their lack of expertise at the task would be masked by the lack of feedback on accuracy.

Our musicians showed greater timbre invariance than nonmusicians. This result suggests a different approach to timbre invariance. Handel and Erickson (2001) treat timbre invariance as a rough constant with their 1-octave rule of thumb. Alternatively, one can approach timbre invariance as an empirically derived function that will change shape depending on musical experience, the instruments under comparison, the note frequencies, and other factors. In this case, it would not be surprising if the timbre invariance function for flute versus piccolo was different from trombone versus piano.

Our results provide partial support for the conclusions of Handel and Erickson (2001). The ability of nonmusicians to discriminate whether two notes were played by the same instrument, or not, declined to chance levels when the octave separation of the notes was increased. In contrast, musicians showed only a slight decline in discrimination accuracy. The difference in performance between musicians and nonmusicians suggests that other factors must be investigated before one can conclude that the bandwidth for timbre invariance is one octave.

Author Note

Address correspondence to: Kenneth M. Steele, Department of Psychology, Appalachian State University, Boone, North Carolina 28608. E-MAIL *steelekm@appstate.edu*

References

ANSI (1973). *Psychoacoustical terminology. S3.20*. NY: American National Standards Institute.

BEAL, A. L. (1985). The skill of recognizing musical structures. *Memory and Cognition, 13*, 405-412.

FURR, R. M., & ROSENTHAL, R. (2003). Repeated-measures contrasts for "multiple-pattern" hypotheses. *Psychological Methods, 8*, 275-293.

HAJDA, J. M., KENDALL, R. A., CARTERETTE, E. C., & HARSHBARGER, M. L. (1997). Methodological issues in timbre research. In I. Deliege & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 253-306). Hove, East Sussex: Psychology Press.

HANDEL, S., & ERICKSON, M. L. (2001). A rule of thumb: The bandwidth for timbre invariance is one octave. *Music Perception, 19*, 121-126.

HART, J. T. (1965). Memory and the feeling of knowing experience. *Journal of Educational Psychology, 56*, 208-216.

IVERSON, P., & KRUMHANSEL, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America, 94*, 2595-2603.

NELSON, T. O., GERLER, D., & NARENS, L. (1984). Accuracy of feeling of knowing for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113*, 282-300.

OPOLKO, F., & WAPNICK, J. (Producers). (1987). *McGill university master samples* [Compact Disk]. Montreal, Quebec: McGill University.

OPOLKO, F., & WAPNICK, J. (1989). *McGill university master samples user's manual*. Montreal, Quebec: McGill University.

PERFECT, T. J., & HOLLINS, T. S. (1999). Feeling-of-knowing judgments do not predict subsequent recognition performance for eyewitness memory. *Journal of Experimental Psychology: Applied, 5*, 250-264.

PITT, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 976-986.

RISSET, J. C., & WESSEL, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 113-169). San Diego, CA: Academic Press.