EFFECTS OF DARK-BRIGHT TIMBRAL VARIATION ON THE PERCEPTION OF FLATNESS AND SHARPNESS

Joel Wapnick
Peter Freeman

Fifty undergraduate music majors listened to 48 pairs of clarinet tones. Both tones in each pair were three seconds long, the interstimulus interval lasted one second, and the interval between trials was ten seconds. The subjects' task was to indicate whether the pitch of the second tone sounded sharp, flat, or the same as the first pitch. The tones had been altered by an audio equalizer so that they were either bright or dark relative to the original recorded level. The equalized tone pairs thus consisted of bright-bright, dark-dark, bright-dark, and dark-bright sequences. In addition, the second tone was adjusted to either 12 cents sharp, 12 cents flat, or was left unaltered. Analyses of response error patterns suggest that subjects associated darkness with flatness and brightness with sharpness.

Results from recent experimental research suggest that the perception of timbre and pitch are interrelated. Most of this research has focused on the effects of timbre on pitch. For example, both pitch discrimination (Henning & Grosberg, 1968; Houtsma, 1971; Sergeant, 1973; Zeitlin, 1964) and pitch matching in music performances (Greer, 1970) have been shown to be superior when the stimuli consisted of timbres characterized by many overtones than when stimuli were of timbres characterized by few or no overtones. It also appears that other types of systematic differences in pitch discrimination ability exist as a function of different instrumental and electronic timbres (Meyer, 1978). In addition, the ability to fine tune pitch through dial manipulations has been shown to be affected by differences in instrumental timbre (Swaffield, 1974).

The perception of timbre apparently is affected by systematic variations in pitch. For example, Saldanha and Corso (1964) found that the percentage of correctly recognized instrumental timbres was significantly affected by pitch level. Madsen and Geringer (1976) demonstrated that subjects' timbral preferences for "good" over "bad" unaccompanied trumpet tone quality could be masked by the addition of another pitch structure, such as an arpeggiated triadic accompaniment.

The purpose of the present investigation was to determine the degree to which timbral variation along a "dark-bright" continuum might affect

Requests for reprints should be sent to Joel Wapnick, Faculty of Music, McGill University, 555 Sherbrooke Street West, Montreal, PQ, Canada H3A 1E3.

the perception of flatness and sharpness. "Dark" and "bright," as used in this study, refer to the relative amplitudes of lower versus higher overtones that comprise clarinet tone. Dark and bright timbres were thus created by using an audio equalizer to alter a previously unbiased recorded clarinet tone. Subjects listened to trials consisting of pairs of tones, with timbral sequences for each pair being either bright-bright, dark-dark, bright-dark, or dark-bright. In addition, the second tone was adjusted sharp or flat, or it remained unchanged in pitch level relative to the first tone. The subjects' task was to indicate whether they thought that the second tone was sharp, flat, or the same as the first.

Method

Subjects were 50 randomly selected undergraduate music majors enrolled in the McGill University Faculty of Music. The experiment was conducted in an acoustically insulated 15 by 25 foot room. Audio equipment used for the presentation of tonal stimuli consisted of a SONY TC-377 tape recorder, two KLH 4 loudspeakers, a Quad 33 preamplifier, and a Quad 303 power amplifier.

A professional clarinetist recorded the tones used to prepare the experimental tape. He performed two concert A pitches (220 Hz and 880 Hz) for three seconds each, at a mezzo forte level. A Korg Tuning Standard (model WT-10) ensured that the recorded tones would be pitched as accurately as possible. The tones were recorded at 15 ips without any equalization. Recording equipment consisted of two AKG D202-E1 microphones and the SONY TC-377 tape recorder. The tape recorder was cleaned and calibrated to avoid artificial coloring of the recordings due to machine maladjustments. Microphone placements were established after testing by a sound recording engineer to avoid artificial coloring. Recordings were made on Scotch 206 magnetic recording tape.

Samples of bright and dark tone for each note were obtained by passing the original two tones through a seven band JVC SEA-20 graphic audio equalizer and adjusting the levers as shown in Table 1. A Heathkit SM

Table 1
Audio Equalizer Settings Used to Create Dark and Bright Clarinet Tones

		Audio Equalizer Bands						
Register	Timbre	60 Hz	150 Hz	400 Hz	1 kHz	2.4 kHz	6 kHz	15 kHz
Low	Dark	2	4	3	0	−6	−8	−10
	Bright	−4	−4	−2	4	6	4	0
High	Dark	0	0	10	7	−8	−10	−12
	Bright	0	0	−6	−2	4	6	0

Note: Numbers in the table indicate changes in decibel levels made to each band in order to produce the altered timbres.

4100 frequency counter was subsequently used to recheck the frequencies of the altered tones. There was no difference in bright-tone and dark-tone frequencies for these tones recorded at A=220 Hz or for these tones recorded at A=880 Hz. Loudness levels for bright and dark tones that were used in the experiment were established in pilot testing that involved six undergraduate music majors; dark and bright tones were adjusted in overall loudness so that they appeared to be equal to these subjects.

Four members of the McGill University woodwind faculty then listened to ten pairs of the equalized tones. They were asked to indicate which instrument or instruments they were listening to, and whether the timbral sequence of each pair was the same, bright-dark, or dark-bright. A 100% accuracy of response rate for both questions was obtained, which indicated that the equalized tones were recognizable as clarinet timbres, and that the timbral variation affected through the use of the equalizer was great enough to be reliably perceived.

An Ampex 440C variable-speed tape recorder was used in conjunction with a Moog synthesizer to raise or lower the pitch of the recorded tones by 12 cents. This amount of pitch deviation was determined through earlier pilot testing that indicated that when timbre and pitch of two tones were varied simultaneously, pitch differences of less than eight cents between the tones resulted in unreliable pitch judgments. Differences of over 15 to 20 cents resulted in pitch judgments that were almost always correct, and thus were not subject to the influence of timbral differences between the two tones. The oscillator in the synthesizer produced a sine tone and controlled the speed of the tape recorder so that when the frequency counter indicated 9,600 Hz, the variable-speed motor was running at precisely 15 ips. Increasing the oscillation so that the counter indicated 9,667 Hz raised the pitch 12 cents, and decreasing the oscillation so that the counter indicated 9,534 Hz lowered the pitch 12 cents. As the pitch of a tone is raised or lowered, the timbre also changes, but small pitch changes required in this experiment were not expected to noticeably alter the clarinet timbres.

Trials of clarinet tone pairs were then constructed. The flatted clarinet tones were 3.02 seconds long, and sharpened clarinet tones were 2.98 seconds long. Flatted tones were thus about 1.3% longer than sharpened tones, which was considerably smaller than the 10% change considered necessary for the durations of two tones to be reliably discriminated (Cogan & Pozzi, 1976; Creelman, 1962). The interstimulus interval was one second, and the intertrial interval was 10 seconds. Four timbral sequences were employed for the tone pairs: dark-dark, dark-bright, bright-bright, and bright-dark. In addition, the second tone was 12 cents sharp, 12 cents flat, or unaltered in pitch, relative to the first tone. This resulted in a total of 24 trials: two notes (A=220 Hz and A=880 Hz) × four timbral sequences × three intonation conditions. An experimental tape consisting of 48 trials was then produced; the second 24 trials were the same as the first 24, and were used to determine intrasubject reliability. Each set of 24 trials was presented in a different random order.

Procedure

Upon entering the experimental setting, subjects were seated at desks facing the two wall-mounted loudspeakers and were given answer sheets. The duration of each experimental session was approximately 20 minutes. Subjects were tested in small groups of 12 to 15 at a time. The following instructions were given orally.

> This is an experiment concerning pitch discrimination. Shortly you will hear a number of paired clarinet tones. Please indicate on the answer sheets before you whether you think the second tone of each pair is sharper than, flatter than, or the same as the first tone. There are 48 trials in all. A three-minute rest period will be provided at the halfway point. We are ready to begin. Are there any questions?

Results

Response errors were grouped into flat and sharp categories. A flat error was defined as one in which the subject indicated that the second tone was flat relative to the first when it was the same as or sharper than the first; or the subject indicated that the second tone was the same as the first when it was sharper than the first. A sharp error was defined as one in which the subject indicated that the second tone was sharp relative to the first when it was the same as or flatter than the first; or the subject indicated that the second tone was the same as the first when it was flatter than the first.

Timbral sequence accuracy

A *t* test for related measures was used to examine subjects' errors as a function of the same versus different timbral sequence trials. As can be seen in Table 2, subjects made significantly more errors in trials involving different timbre (dark-bright and bright-dark) than they did in trials involving no timbral change (dark-dark and bright-bright), regardless of whether low tones or high tones were heard.

Next, sharp versus flat errors were examined as a function of timbral sequence. *t* tests were performed on each of the four possible sequences. Subjects made significantly more flat errors than sharp errors when the timbral sequence of trials was bright-dark (see Table 3). When the timbral sequence was dark-bright, subjects made significantly more sharp errors than flat errors. No difference was found between the number of sharp and flat errors when the timbral sequence was bright-bright. When the timbral sequence was dark-dark, however, subjects made significantly more flat errors than sharp errors.

Two other tests determined the degree to which subjects made their pitch judgments on the basis of the second tone alone versus the degree to which they made their judgments on the basis of the timbre of the second tone in relation to the timbre of the first. A *t* test was performed on flat errors made on trials involving bright-dark timbre as compared to trials involving dark-dark timbre. A *t* test was also performed on sharp errors made on trials involving dark-bright timbre as compared to trials involving bright-bright timbre. In both cases subjects made significantly more errors when the timbre changed as compared to when it remained the same (bright-bright versus dark-dark: $t=7.17$, $df=49$, $p < .001$; dark-bright versus bright-bright: $t=8.48$, $df=49$, $p < .001$).

Accuracy of judgments as a function of intonation

A one-way repeated measures analysis of variance was performed to study subjects' accuracy as a function of whether the second tone of each trial was flat, the same, or sharp relative to the first tone in the trial (see Table 4). There was an overall main effect ($F=9.28$, $df=2, 49$, $p<.001$).

Table 2
Average Errors Per Subject as a Function of Same Versus Different Timbral Sequences

	Same Timbral Sequence (bright-bright and dark-dark)	Different Timbral Sequence (bright-dark and dark-bright)	*t*	*df*	*p*
Total Mean Errors Per Subject	8.80	11.68	6.41	49	<.001
Standard Deviation	3.31	2.92			
Average Errors Per Subject for Low Frequency Tones (A=220 Hz)	4.68	6.10	4.60	49	<.001
Standard Deviation	1.80	1.71			
Average Errors Per Subject for High Frequency Tones (A=880 Hz)	4.12	5.58	5.59	49	<.001
Standard Deviation	2.09	1.85			

Table 3
Average Sharp Versus Flat Errors Per Subject as a Function of Timbral Sequence

Timbral Sequence	Sharp Errors	Flat Errors	*t*	*df*	*p*
Bright-Bright	2.30	2.28	0.07	49	NS
Dark-Dark	1.46	2.76	3.99	49	<.001
Bright-Dark	0.90	5.04	10.87	49	<.001
Dark-Bright	4.70	1.04	9.84	49	<.001

The Newman-Keuls procedure was then employed to determine differences among the three intonation condition means. The mean number of errors per subject when the second tone was sharp (7.04) was greater than the mean number of errors per subject when the second tone was flat (5.66). The mean number of errors per subject when the second tone was the same as the first (7.78) was also greater than that found when the second tone was flat ($p < .01$ for both comparisons). No significant difference was found between the mean number of errors per subject when the second tone was sharp versus when it was the same as the first. Subjects' responses were apparently more accurate when the second tone of each trial was flatter than the first, as compared to when it was the same as or sharper than the first.

Reliability

Intrasubject reliability was determined by dividing the number of agreements by agreements plus disagreements, and then multiplying the result by 100 to obtain a percentage score. An agreement was defined as an instance in which a subject gave the same response to any given trial occurring in the first half of the experiment and its replication in the second half. With three responses — flat, same, and sharp — chance agreement was 33.3%. Intrasubject reliability was 62%. A *t* test was calculated in order to compare the average number of errors per subject (20.48) with the chance rate of 32 errors out of 48 trials. Subjects made significantly fewer errors than that expected through chance responding ($t < 21.8$, $df = 47$, $p < .001$).

A Pearson product-moment correlation coefficient was calculated to determine intersubject split-half reliability. The two scores for each subject used for calculating this correlation consisted of the numbers of errors made in odd-numbered trials and the number of errors made in even-numbered trials. The correlation coefficient obtained in this manner was .62.

Table 4
Analysis of Variance for Errors as a Function of the Intonation of the Second Tone Relative to That of the First

Source	*SS*	*df*	*MS*	*F*	*p*
Total	1198.49	149			
Subjects	471.49	49			
Treatment (Intonation)	115.77	2	57.89	9.28	<.001
Error	611.23	98	6.24		

Discussion

The statistically significant greater number of errors made in trials involving a change in timbre as opposed to no change in timbre suggests that tone quality is an important factor in pitch discrimination. Subjects often associated changes in timbre with changes in pitch. When direction of error was examined, subjects apparently associated darkness (when preceded by brightness) with flatness, and brightness (when preceded by darkness) with sharpness. It would be interesting to determine whether these effects occur in music performance situations. If so, a portion of inaccuracy that has been found in studies of music performance intonation (Mason, 1960; Nickerson, 1949; Schoen, 1922; Small, 1937) may have been due to timbral fluctuations affecting pitch perception.

The fact that response accuracy was systematically affected by timbral variation does not necessarily mean that incorrect responses were "wrong" from a musical standpoint. Results from the experiment show considerable agreement across subjects concerning the way in which timbral variation affected the directionality of incorrect judgments. This implies that technically inaccurate judgments were socially acceptable, and therefore could be considered to have been "good" judgments.

Although there were no differences between the average number of flat and sharp errors per subject for bright-bright trials, more flat than sharp errors were made for dark-dark trials. Subjects had apparently made their pitch judgments within dark-dark trials on the basis of the second tone alone. To account for this, it was hypothesized that bright tone used in this experiment may have sounded less like a clarinet than dark tone. In boosting the lower equalizer bands and dropping the higher equalizer bands to make the original tone darker, the lower and predominantly odd-numbered overtones were increased in loudness whereas the higher and odd- and even-ordered overtones were decreased in loudness. This procedure may have augmented the clarinet-like tone characteristics for which the instrument is recognized. However, in making the original tone brighter the lower equalizer bands were dropped and the higher bands were boosted, which decreased the loudness of the lower and predominantly odd-numbered overtones and increased the loudness of the higher and odd- and even-numbered overtones. The terms bright and dark describe tones that are bright or dark relative to a standard instrumental sound or voice quality. If the bright timbre employed in this experiment was less clarinet-like in quality than the timbre of a typical clarinet, then the term bright as it applies here would have little relevance. This semantic difference would explain why subjects might have associated flatness with the darkness of the second tone in dark-dark sequences, but did not associate sharpness with the brightness of the second tone in bright-bright sequences.

Fewer errors were made when the second tone was flat relative to the first tone than when the second tone was the same or sharper. This is consistent with other research showing that the perception of flatness is generally keener than the perception of sharpness for musicians (Madsen, Edmonson, & Madsen, 1969; Madsen & Geringer, 1976; Siegel & Siegel, 1977).

Since this study was concerned with intonation perception rather than intonation performance, it would be improper to imply from it that training of any particular kind might be effective in improving intonation performance. Moreover, Geringer's (1978) finding that intonation perception of ascending scales was different (sharper and less accurate) than performed intonation of these scales suggests that the two types of intonation may not be closely related.

The acquisition of intonation listening skills appears to be a desirable goal for many musicians. Such skills are especially crucial in teaching situations where it is important to give students who sing or play nonfixed pitched instruments accurate feedback concerning their intonation. The relevance of the present study to applied music instruction is contained in its demonstration of how timbre may affect intonation perception. Although the study does not suggest specific training methods for learning accurate intonation perception, it does imply that in such an approach it would be important to be aware of timbral effects. Further research might be helpful in determining if the effects found in the present experiment occur under a variety of different stimulus conditions (for example, different instruments, different note durations, and melodies as well as isolated tones) and in developing techniques for teaching intonation perception in a manner that incorporates the study of intonation as a function of timbral variation.

References

Cogan, R., & Pozzi, E. *Sonic design: The nature of sound and music.* Englewood Cliffs, N.J.: Prentice-Hall, 1976.

Creelman, C.D. Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, 1962, *34*, 582-593.

Geringer, J.M. Intonational performance and perception of ascending scales. *Journal of Research in Music Education,* 1978, *26,* 32-40.

Greer, R.D. The effect of timbre on brass-wind intonation. *University of Iowa Studies in the Psychology of Music,* 1970, *6,* 65-94.

Henning, G.B., & Grosberg, S.L. Effects of harmonic components on frequency discrimination. *Journal of the Acoustical Society of America*, 1968, *44*, 1,386-1,389.

Houtsma, A.J. What determines musical pitch? *Journal of Music Theory,* 1971, *15,* 138-157.

Madsen, C.K., Edmonson, F.A., III, & Madsen, C.H., Jr. Modulated frequency distribution in relationship to age and musical training. *Journal of the Acoustical Society of America,* 1969, *46,* 1,468-1,471.

Madsen, C.K., & Geringer, J.M. Preferences for trumpet tone quality versus intonation. *Bulletin of the Council for Research in Music Education*, 1976, *46*, 13-22.

Mason, J.A. Comparison of solo and ensemble performances with reference to Pythagorean, just, and equi-tempered intonation. *Journal of Research in Music Education,* 1960, *8,* 31-38.

Meyer, J. The dependence of pitch on harmonic sound spectra. *Psychology of Music,* 1978, *6,* 3-12.

Nickerson, J.F. Intonation of solo and ensemble performance of the same melody. *Journal of the Acoustical Society of America,* 1949, *21,* 593-595.
Saldanha, E.L., & Corso, J.F. Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America,* 1964, *36,* 2,021-2,026.
Schoen, M. An experimental study of the pitch factor in artistic singing. *Psychological Monographs,* 1922, *31,* 230-259.
Sergeant, D. Measurement of pitch discrimination. *Journal of Research in Music Education,* 1973, *21,* 3-19.
Siegel, J.A., & Siegel, W. Categorical perception of tonal intervals: Musicians can't tell flat from sharp. *Perception & Psychophysics,* 1977, *21,* 399-407.
Small, A.M. An objective analysis of artistic violin performance. *University of Iowa Studies in the Psychology of Music,* 1937, *4,* 172-231.
Swaffield, W.R. Effect of melodic parameters on ability to make fine-tuning responses in context. *Journal of Research in Music Education,* 1974, *22,* 305-312.
Zeitlin, L.R. Frequency discrimination of pure and complex tones. *Journal of the Acoustical Society of America,* 1964, *36,* 1,027.

■ McGill University
Montreal, Canada
October 3, 1979