sempre:
Society for Education, Music and Psychology Research

Article

Psychology of Music
39(3) 291–306

DOI: 10.1177/0305735610373602
pom.sagepub.com
SAGE

Does timbre affect pitch?: Estimations by musicians and non-musicians

Allan Vurma, Marju Raju and Annika Kuuda
Estonian Academy of Music and Theatre, Estonia

Abstract

The present article focuses on the question of whether the timbre difference of two sounds with harmonic spectra, produced by natural musical instruments or the singing voice, may influence subjective assessments of the pitch of one sound in relation to the pitch of the other. The authors administered a series of perception tests to a group of professional musicians (n = 13) and a group of non-musicians (n = 13). The tests used the following pre-recorded sounds: the singing voice, the sound of the viola, and the sound of the trumpet. The participants had to compare the pitch of pair-wise presented successive tones and decide whether the second tone was either 'flat', 'sharp' or 'in tune'. Tests using stimuli in the pitch range around A3 (220 Hz) at a loudness level of approximately 90 phons revealed pitch shifts of significant magnitude likely to affect intonation quality in a musical performance among both musicians and non-musicians. The conclusion drawn from the study is that timbre-induced pitch shifts may attain magnitudes that are likely to lead to conflicts between subjective and fundamental-frequency-based pitch assessments. Situations are described in which such conflicts may arise in actual musical practice.

Keywords

pitch, pitch perception, pitch shift, timbre, tuning

Introduction

The pitch of a sound is often identified as its fundamental frequency or the frequency of vibration, e.g., the A4 is worldwide set to 440 Hz. Still, other attributes of the sound, such as its sound pressure level (SPL), spectral composition, duration, and the simultaneous occurrence of other sounds are found to affect perceived pitch (Terhardt, 1988). For example, it has been demonstrated (e.g., Moore, Glasberg, & Peters, 1985) that when a spectral component in a sound with harmonic spectrum is mistuned by up to 8%, the perceived pitch of that sound shifts in the direction of mistuning, and that the pitch of sine tones below 1000 Hz appears lower and above 2000 Hz higher when the SPL of the sound increases (Stevens, 1935). Therefore, the American National Standards Institute defines pitch as an auditory attribute of

Corresponding author:
Allan Vurma, Estonian Academy of Music and Theatre, Rävala 16, Tallinn 10143, Estonia.
[email: vurma@ema.edu.ee]

sounds that allows sounds to be ordered on a scale from low to high but does not represent an objectively measurable parameter (ANSI, 1994).

The timbre of a sound is that attribute of auditory sensation that allows a listener to differentiate between two sounds that are presented in a similar manner and have the same loudness and pitch (ASA, 1960). Traditionally, timbre is linked to the distribution of energy in the power spectrum; still, many other features, like the temporal patterning of the parameters and the presence of different noise components, also contribute to the perception of timbre (Moore, 2003).

The influence of timbre on pitch perception in the case of sounds with harmonic spectrum is less obvious. For example, in the experiment of Chuang and Wang (1978) with synthesized vowels, by using the multivariate design procedure of pair-wise comparison of stimuli, the interactive factors (vowel quality difference, intensity difference, and F0 difference) for vowel pitch perception were studied. The vowel quality difference (but not the intensity difference) produced significant pitch interaction: the vowel /a/ with F0 at 100 Hz was perceived as 0.54, 1.25 and 2.8 Hz higher in pitch in comparison to vowels /ɛ/, /i/, and /u/ with the same F0. Later, the results of experiments conducted by Pape and Mooshammer (2006) with natural speech sounds showed that this so-called intrinsic pitch of vowels could be language- and expertise-dependent. The phenomenon was present in German listeners and vowels but was not noticeable in Italian listeners either with Italian or with German vowels. Also, the German musicians did not show any significant shift in perceived pitch caused by the vowel quality difference.

In other investigations (e.g., Singh & Hirsh, 1992; Warrier & Zatorre, 2002) the interaction between timbre and pitch was still observed, but was done so by using synthetic sounds quite unlike those of natural musical instruments and in more remote experimental settings than the direct comparison of pitches common in music practice. For example, Singh and Hirsh (1992) compared specially synthesized harmonic complex tones whose partials differed in their amplitude from zero in only one single frequency band. They discovered that shifting a locus of that band on the frequency axis was likely to have a stronger impact on the perception of pitch than small changes in the fundamental frequency of the sound. Warrier and Zatorre's (2002) experiment focused on comparing the pitch of harmonic complex tones of 'low', 'middle' and 'high' timbre. All tones consisted of 11 partials. The levels of consecutive partials were either monotonically increasing, monotonically decreasing, or increasing for partials two to six and decreasing for partials seven to eleven. The intensity of the fundamental was always kept constant. The fundamental frequency of the tones was manipulated upwards in steps of 0, 17, 35 and 52 cents. The experiment showed that in all cases a change in timbre also caused the listener to perceive the pitch differently, although the difference was smaller when test tones were presented as part of a melody. In a recent experiment conducted by Vurma and Ross (2007), classically trained singers were asked to reproduce the pitch of synthesized sounds having the timbre of either the piano or the oboe. On average, the fundamental frequency of singing voice reproductions of synthesized stimuli was 13 cents (piano timbre) and 7 cents (oboe timbre) lower than the respective stimulus. In their second experiment, the consecutive sounds of operatic tenor voice with vibrato and the synthesized sounds resembling piano or oboe were presented pair-wise to the musically educated listeners. They were asked whether the second tone in the pair was 'flat', 'sharp', or 'in tune'. The majority of 'in tune' results corresponded to the piano or oboe sounds that had a 15–20 cents higher F0 than the average F0 of the tenor's voice. Still, it was not clear whether the unevenness of wide vibrato of the singing voice (and the phase distortions of its manipulated version with straightened vibrato) could influence the results.

The present study avoids the use of synthesized sounds and focuses on the following questions: (1) In a comparison of two sounds with harmonic spectra produced by natural musical instruments or the singing voice, similar to those carried out by a musician in a performance setting, is the timbre difference of those sounds likely to cause a shift in listeners' pitch relationship judgments (between 'in tune', 'flat', and 'sharp') in a systematic manner and with a magnitude that is significant in terms of the intonation quality of musical performance; and (2) Are the characteristics of such a shift dependent upon whether or not the listeners possess a musical education? Given that timbral influences exist, pitch judgments may not match precisely the actual values of F0. Since the nature of pitch assessments is probabilistic, it would be of considerable interest to musicians to learn about, in addition to the likely magnitudes of timbre-induced pitch shifts, related changes in the probability of the three alternatives. From a practical perspective, it would also be important to present such information with reference to examples of sounds produced by musical instruments that musicians use in their practice.

Methods

Participants

1. Test group I: Professional musicians. Test group I consisted of 13 individuals (eight men and five women). Ten were students at the Estonian Academy of Music and Theatre, representing different specialities (piano, oboe, trumpet, violin, cello, conducting, and classical singing). The average age of the students was 22.5 years. The remaining three participants in test group I were professional musicians (two choir conductors and one singer) between 45 and 50 years of age. All individuals in the group had started to study music at an early age and continued to engage in music activities as part of their professional lives on a daily basis. Two participating pianists were also active as choir singers (i.e., their work as musicians included situations where they had to intone pitches).

2. Test group II: Non-musicians. Test group II also consisted of 13 individuals (four men and nine women). Nine of the participants were between 23 and 27 years of age, one was a 14-year-old and the remaining three were 34, 39, and 64 years of age. Seven had learned a musical instrument (such as the piano, accordion, recorder, or guitar) for a few years as a child. None of the participants of group II had actively engaged in any musical activities up to the time of the test, i.e., music was neither their profession nor a hobby. Test group II resembles ordinary concert-goers – people who need not be professional musicians, but who may have studied a musical instrument during their primary schooling. According to the information provided by the participants, no one in either group suffered from any hearing impairment (although this was not checked).

Stimuli

The sound stimuli were based on recordings of the note A3 (fundamental frequency F0 = 220 Hz). The recorded sounds were (1) produced on the viola using a bow, (2) played on the trumpet, or (3) sung by a classically trained tenor as the vowel /a/. The selection was based on (1) a clear contrast between the timbres of the corresponding sounds, (2) a partial overlap of the pitch ranges of the viola, the trumpet, and the tenor voice, which allowed us to select a sound of the same F0 from each of these, and (3) the fact that the spectra of the sounds of all three sources are harmonic. The instruments were recorded in the Anechoic Chamber of the

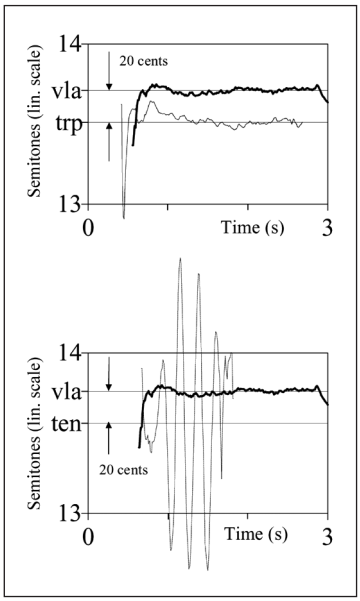

Figure 1. F0 fluctuations of sounds used in perception tests

Notes: Abscissa: time (seconds); ordinate: relative F0 (semitones). The average F0 shift between viola and trumpet (upper panel) and between viola and tenor voice (lower panel) is 20 cents. Regardless of the shift, the participants perceived the corresponding pairs of sounds to reflect the best match in pitch. The distance between arbitrary values 13 and 14 depicted in the graphs corresponds to an F0 difference of one semitone.

Wendell Johnson Speech and Hearing Center at the University of Iowa.[1] The tenor voice was recorded using an AKG C420 head microphone (the distance between the microphone and the corner of the singer's mouth was approximately three centimeters) and a SONY TCD10 DAT recorder at the rehearsal studio of Tallinn Philharmonic Society. All sounds were recorded with a sampling frequency of 44.1 kHz. Since we used previously recorded sound samples, we had to accept slight variations in their duration. The duration of the viola and trumpet samples amounted to approximately two seconds and the duration of the voice sample was approximately one second.

Both the viola and trumpet sounds lacked vibrato. As is characteristic of natural musical instruments, the F0 of both sounds exhibited minor random fluctuations within a range of a few cents (Fig. 1, upper panel). The singing voice stimulus was characterized by vibrato, whose amplitude was about 100 cents and frequency approximately 5.7 Hz (Fig. 1, lower panel). Both values are typical for the parameters of the voice of a classically trained opera singer (Sundberg, 1994).

In the spectrum of the viola sound, the energy decreases with increasing frequency (Fig. 2, top panel). In the spectrum of the trumpet sound, the energy of the harmonics increases up to the sixth and decreases after that (Fig. 2, middle panel). The spectrum of the tenor voice is characterized by an energy peak at the third harmonic, a significant drop in energy in the 1.5 kHz range, and a second peak (the so-called singer's formant) in the 2.7 kHz range (Fig. 2, bottom panel). The spectra can also be compared on the basis of the location of their center of gravity, which indicates the average frequency that carries the energy of the corresponding sound. The

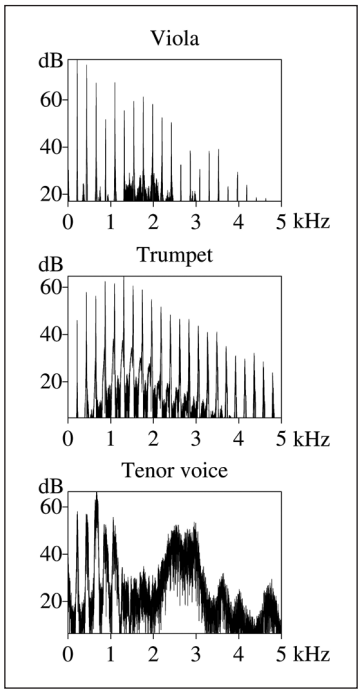

Figure 2. FFT spectra of sounds used in the perception tests

value of the center of gravity (calculated using the Praat 4 software, weighting by absolute spectrum) was lowest (1053 Hz) for the viola sound and highest (1870 Hz) for the tenor voice. The center of gravity of the trumpet sound (1685 Hz) was slightly lower than that of the tenor voice. Thus, sounds with the timbre of the trumpet and the tenor voice can be regarded as brighter than those of the viola.

At the next step, we derived a series of stimuli with different F0s from each original stimulus. Using the pitch correction tool of the professional sound editing software WaveLab 5 (Steinberg Media Technologies GmbH), we increased or decreased the F0 in steps of 5 cents up or down to 50 cents from the value for the original stimulus.[2]

The F0s were described using the average value of the F0 measured over the duration of the quasi-stationary part of the stimulus. The autocorrelation method and the software Praat 4 were used to measure F0 values.

Description of perception tests

Participants were played pairs of sounds (reference sound followed by test sound) and were asked to indicate whether the second sound was in tune, flat, or sharp compared to the first. Such a test paradigm is similar to real-life situations, where a performing musician first makes the decision and only then adjusts the pitch if he/she feels it necessary. The sounds were played binaurally to the participants via Sennheiser EH1430 earphones using a computer equipped with a sound card. In subtests, the first (reference) sound (trumpet, viola, or singing voice) in the pair always had the same F0. The second sound, depending on the subtest, was either of the

same timbre as the first sound or of a different timbre. The F0 of the second sound was either the same as that of the first sound or differed from it by up to 50 cents up or down, depending on the particular modification used in the pair. The two sounds were separated by approximately two seconds of silence. The following combinations were represented in the six tests: viola–viola (test 1), trumpet–trumpet (test 2), viola–trumpet (test 3), trumpet–viola (test 4), viola–tenor voice (test 5), tenor voice–viola (test 6). In a single test, nine modifications of the test sound with different F0s were used. The sound pairs with different pitch modifications were played to the participants in a random order that differed each time. Each pitch modification was played five times during a test. Thus, participants were exposed to $9 \times 5 = 45$ pairs of sounds in a test. The order of the tests was random and different for each participant.

In designing and preparing the experiment we were trying not to overload the participants and to keep the number of stimulus-pairs small. We were attempting not to include the ranges of stimuli which would have only been rated with high probability either 'flat' or 'sharp'. For that purpose we conducted a limited pilot test to predict the approximate range of pitch shift for all timbre pairs involved in the study. The results of pilot tests with stimuli that were distributed from minus 50 to plus 50 cents symmetrically around the zero F0 deviation point showed clearly the asymmetry of responses. Therefore we decided in the main tests to use a more narrow and asymmetrical F0 deviations area. It is possible that if during the test only one type of response (e.g., 'flat') prevailed, the participant, if hesitant, would start to subconsciously counterbalance this with more frequent opposite (i.e., 'sharp') responses, which would finally lead to the bias of the results.

The results of the preliminary test with singing voice showed noticeable wider 'in tune' response curves in comparison with the other tests. To ensure the delineation of corresponding graphs without increasing the number of stimulus-pairs, we increased the F0 increment of test sounds with the factor of two in these tests.

The loudness levels of all reference and stimulus sounds were adjusted to be equal (ca. 90 phons) by a professional musician (by using a sine tone with F0 = 1000 Hz and SPL = 90 dB as a reference). The measured SPL (using B&K head and torso simulator type 4128C, flat frequency response) of the various stimuli fluctuated less than 2 dB. The purpose of the high SPL in our tests was to imitate real sound volumes experienced by musicians on the stage (as opposed to those to which listeners in the audience are exposed).

We used the modified perception tests module of the software Praat 4 to administer the test to the participants. The participants had to use the mouse to click on one of the three squares displayed on a computer screen to indicate their assessment. As an alternative, they could also use the keyboard. While it was not possible for the participants to replay a test pair, the time they had for making up their mind about how to assess the pair was not limited. Each subsequent pair of sounds was played approximately one second after the participant had indicated his or her assessment of the previous pair. After each tenth assessment, participants were allowed to take a break of unlimited duration. In each subtest, participants could at any time display information about the number of pairs they had already assessed and the number of pairs remaining in the subtest. Depending on the participant's speed of making assessments and the time spent on breaks, the six tests took from 25 to 60 minutes to complete.

Results

Figures 3, 4, and 5 plot the distribution of in tune assessments (top panels), and both flat and sharp assessments (bottom panels) depending on the F0 shift of the test sound for test group I (professional musicians) and test group II (non-musicians).

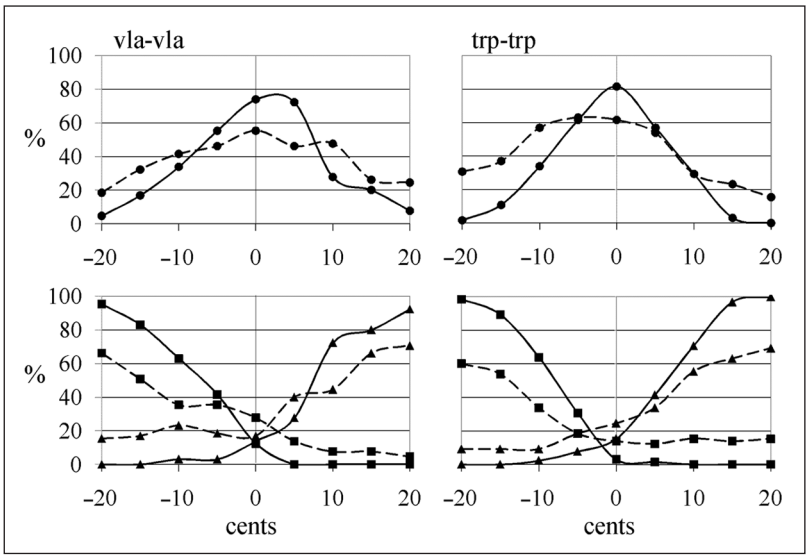

Figure 3. Results of viola–viola test (left) and trumpet–trumpet test (right)

Notes: Upper panels depict the distribution of 'in tune' ratings, lower panels the distribution of 'flat' (rectangles) and 'sharp' (triangles) ratings. Continuous lines show the ratings of musicians, dashed lines the ratings of non-musicians. Abscissa: F0(stimulus)–F0(reference) in cents; ordinate: percentage of ratings.

Tests involving sounds of the same timbre: Viola–viola (1) and trumpet–trumpet (2)

As predicted, the distribution of 'in tune' ratings in tests that involved pairs of sounds of the same timbre (Fig. 3) was close to normal (slightly less so in the case of the test group of non-musicians) and the distribution curves for 'sharp' and 'flat' ratings, respectively, either rose or fell monotonically. The peak of 'in tune' ratings and the intersection point of the curves of 'flat' and 'sharp' ratings are located near zero shift.

Some investigations have shown the bias of musicians to notice and correct flat intonation more frequently than sharp intonation (e.g., Salzberg, 1980). The results shown in Figure 3 indicate that the effect of a bias of this type on the assessments was insignificant. When compared to those of musicians, the non-musicians' distribution curves of 'in tune' ratings were more flat, their curves peaked at lower percentage values, and their differentiation between flat and sharp intonation was poorer. In the area within 15 cents up and down from the reference, the difference between the distribution curves of 'in tune' ratings of musicians and non-musicians was statistically significant in both trumpet–trumpet ($\chi^2 = 21.6(6)$, $p = .001$) and viola–viola ($\chi^2 = 13.6(6)$, $p = .03$) tests.

It is noteworthy that in approximately 20% and 40% of cases respectively, when a sound pair consisted of two identical sounds, musicians and non-musicians did not rate the pair as 'in tune'. At the same time, musicians very rarely failed to determine the direction of frequency modification accurately, even when the modification was only 5 cents.

From the individual response distributions of musicians it was possible to estimate the intra-group variability of responses (see Table 1). (We will not present corresponding results for non-musicians because of the greater irregularity of their responses.) The individual 'in tune' response peaks location varied between minus and plus seven cents with standard deviation

Table 1. Estimates of average pitch shift and F0 DLs based on results of different tests

	Musicians							Non-musicians
Subtest	Pitch shift	*SD*	Min	Max	DL(individual)	*SD*	DL(group)	DL(group)
vla-vla	0	3	−7	7	11	5	12	22
trp-trp	0	3	−7	7	8	4	12	27
vla-trp	−15	4	−5	−20	10	4	13	25
trp-vla	17	7	5	28	13	4	19	21
vla-ten	−19	11	0	−40	25	7	32	48
ten-vla	17	14	−2	40	18	4	24	52

Notes. All values are given in cents. Pitch shift: the average of the values of individuals; *SD*: standard deviation; Min and Max: smallest and largest values of individual pitch shift; DL(individual): average F0 difference limen of individuals; DL(group): F0 difference limen of the group response. (trp: trumpet, vla: viola, ten: tenor voice). For non-musicians only the group DLs are presented.

(*SD*) of 3 cents in the case of the viola–viola test as well as in the case of the trumpet–trumpet test. The average 'in tune' peaks location across all individuals was at zero shift in both tests.

On the basis of response distributions we also determined the values of F0 difference limens (DL). Our data allow the use of two approaches. The first characterizes individual perception and is determined from individual response distributions. The second approach is that of a musician who gets random estimations about his/her intonation correctness from the audience, which consists of numerous people with their idiosyncrasies of perception. The second type (group DL) can be determined from the group response distributions.

We established frequency deviation values at 75% 'flat' and 75% 'sharp'. We assumed that the peak of 'in tune' ratings was located in the middle of the interval between those values, and the value of the F0 difference limen for sound pairs with a given timbre at a transition point of 75% 'flat' or 'sharp' ratings was therefore half the value of the respective interval. The average F0 DL of individuals was 11 cents (*SD* = 5 cents) in the viola–viola test and 8 cents (*SD* = 4 cents) in the trumpet–trumpet test. These values are close to the 10 cents reported by Sundberg (1991).

We determined the group DLs for the musicians group as well as for the non-musicians group. The group DLs were 12 cents for both the viola–viola and trumpet–trumpet tests in the musicians group, but about two times higher in the non-musicians group: 22 cents for the viola–viola test and 27 cents for the trumpet–trumpet test. The value of the group DL tends to be higher than the average of individual DLs because of the idiosyncratic differences in the location of the centre of individual response distribution curves.

Tests involving sounds of different timbre: Viola–trumpet (3), trumpet–viola (4), viola–tenor voice (5), tenor voice–viola (6)

In the case of tests involving sounds of different timbre, the peak of 'in tune' ratings and the intersection point of the distribution curves of 'flat' and 'sharp' ratings both shifted by approximately 15 to 20 cents from the zero point (Figs. 4 and 5). This shift was observed for both musicians and non-musicians. (The shift was similarly present for those non-musicians who had never taken any music lessons as for those who had learned a musical instrument for a few years.) When participants compared test stimuli of brighter timbre (the trumpet or tenor voice) to reference sounds of duller timbre (the viola), i.e., in the viola–trumpet and viola–tenor voice tests, the maximum number of 'in tune' ratings was given for test stimuli whose F0 was

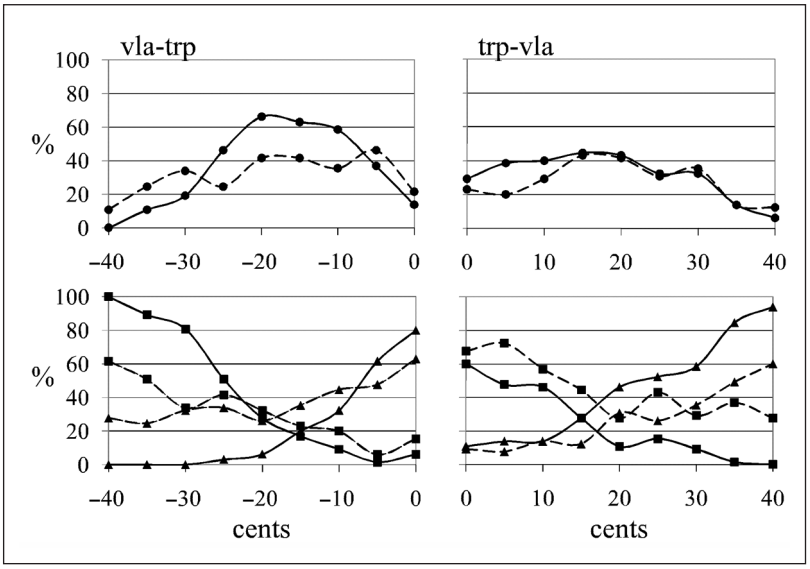

Figure 4. Results of viola–trumpet test (left) and trumpet–viola test (right)

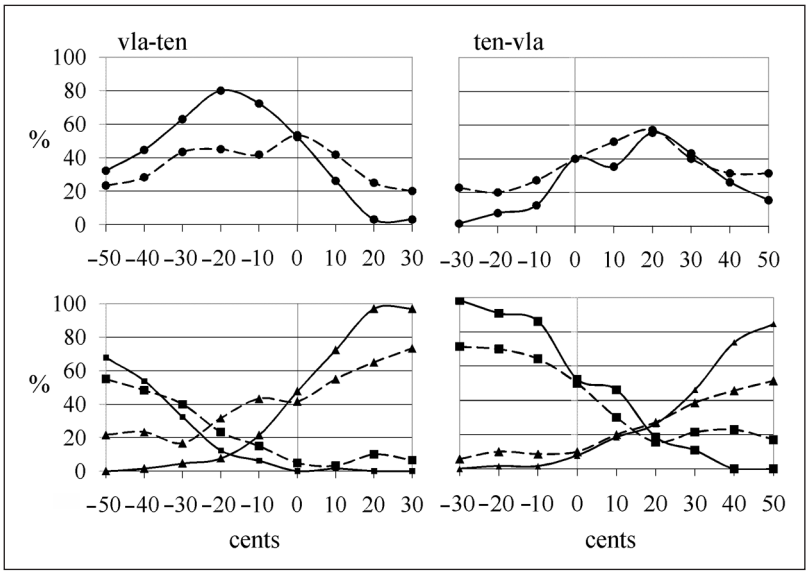

Figure 5. Results of viola–tenor voice test (left) and tenor voice–viola test (right).

approximately 15 to 20 cents lower than the F0 of the reference sound. The same holds true for the reverse situation: in the trumpet–viola and tenor voice–viola tests (duller test and brighter reference sounds) the maximum number of 'in tune' ratings was given for test stimuli whose F0 was approximately 15 to 20 cents higher than the F0 of the reference sound. This shift of the distribution curves along the frequency axis can be interpreted as a timbre-induced shift in perceived pitch.

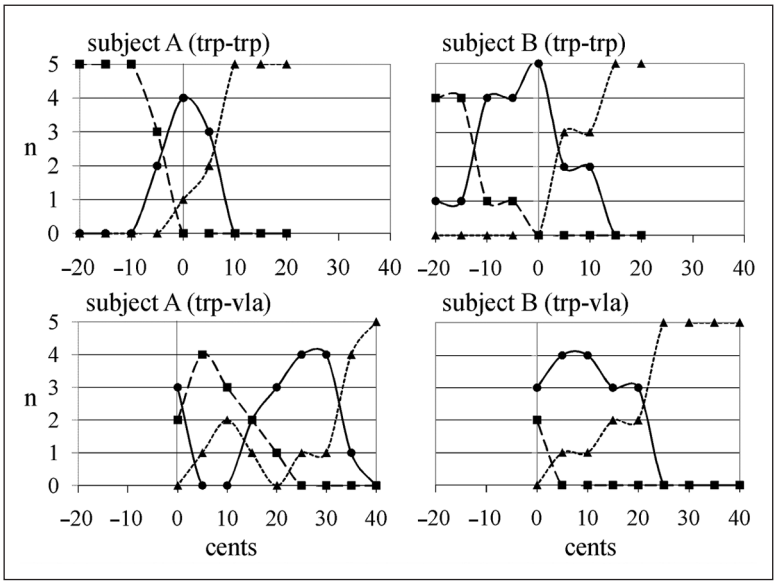

Figure 6. The individual differences of pitch shift

Notes: Left panels show the distribution of 'in tune', 'flat', and 'sharp' ratings given by musician A; right panels show the distribution of corresponding ratings by musician B. Upper panels: trp–trp test; lower panels: trp–vla test.

We used the chi-square test to see whether the change in the proportion of 'in tune', 'flat', and 'sharp' ratings was statistically significant in the tests with different timbres when compared to corresponding responses in the tests with identical timbres. The comparison was performed separately for two frequency deviations. We chose zero frequency deviation as the first locus (it elicited the maximum number of 'in tune' ratings in tests involving sounds with the same timbre) and the frequency deviation of 20 cents from zero manipulation point as the second locus (these had the highest number of 'in tune' ratings in tests involving sounds of different timbre). The difference between respective ratings in both groups was statistically significant (the corresponding χ^2 values lay between 15 and 99, $df = 2$, $p < .001$).

Compared to subtests that involved sounds with the same timbre, the distribution curve of 'in tune' ratings of different-timbre subtests was more uneven and jagged, and had a lower peak. Thus, for example, in the musicians' group, 'in tune' ratings for a single stimulus peaked at 80% in the trumpet–trumpet subtest, but only at 45% in the trumpet–viola test. The difference in the shape of the distribution curve appears to be caused by individual variations in the magnitude of shifts in perceived pitch and by bigger irregularities of individual responses. To illustrate the hypothesis, we have presented a comparison of the distribution of ratings for two musicians (A and B; Fig. 6) in the trumpet–trumpet test (same timbre) and the trumpet–viola test (different timbres). For the trumpet–trumpet test, both participants gave the maximum number of 'in tune' ratings for the sound pair in which the F0 of the test and the reference sounds were the same, whereas in the trumpet–viola test the maximum of the corresponding ratings was located at 25–30 cents (participant A) and at 5–10 cents (participant B). Participant B's relatively large proportion of 'in tune' ratings at the actual zero manipulation point may

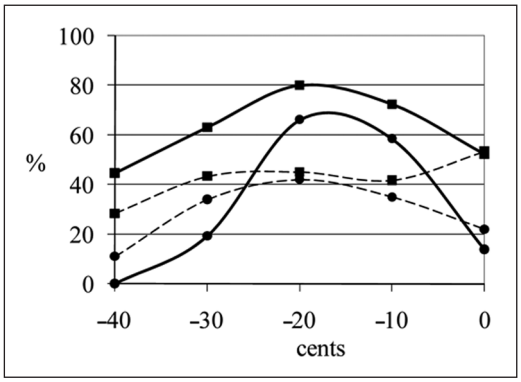

Figure 7. Comparison of the distribution of 'in tune' ratings for viola–tenor voice sound pairs (rectangles) and viola–trumpet sound pairs (circles).

Note: Continuous lines show the ratings of musicians, dashed lines the ratings of non-musicians. Abscissa: F0(stimulus)–F0(reference), in cents; ordinate: percentage of ratings.

indicate that, in some instances, different types of listening may be triggered in the case of sound pairs with different frequency deviations.

In tests where the tenor voice was used, the distribution of 'in tune' ratings extended over a slightly wider range than for the other subtests. Figure 7 shows a comparison of the distribution of 'in tune' ratings in the viola–trumpet and viola–tenor voice tests. The peak of both distribution curves is located at a frequency modification value that is slightly less than 20 cents down from zero. However, at a deviation of –40 cents, approximately 45% of responses in the musicians' group considered the tenor voice as 'in tune', whereas the trumpet sound was universally rated as 'flat'. Tenor voice sounds differed from instrumental sounds in terms of their extensive vibrato and the singer's formant (energy boost at the frequency of 2.6 kHz in the spectrum). Several studies (e.g., Sundberg, 1978; Shonle & Horan, 1980) conclude that the pitch of a tone with vibrato corresponds to the mean F0 of the sound. In those studies, participants were asked to tune sine sounds to match the pitch of a sound with vibrato. The results of our study, which instead of tuning was based on the single ratings, also indicate that when the stimulus is provided by an opera singer using an extensive vibrato (which can mask even a strong random fluctuation in the parameters of the sound) in addition to the shift in perceived pitch, 'in tune' ratings may be 'spread out' over a wider frequency deviation area.

In the tests with different timbres, the individual response distributions of some participants were less regular. In these cases we presumed that the actual 'in tune' peak was located at the imaginary axis of symmetry of the available response distribution curves. (Here we present the data for the musicians only.) There was no difference between the average value of individual pitch shifts and the location of the group 'in tune' response peak (see Table 1 and Figs. 4–5). The intragroup variability was slightly bigger in the tests with different instrumental timbres in comparison to the tests with similar timbres. For example, the individual pitch shift values spanned over 14 cents in the case of viola–viola and trumpet–trumpet tests, but over 23 cents in the trumpet–viola test. The *SD* of individual pitch shift values increased from 3 cents in the tests with similar timbres to 4 cents in the viola–trumpet test and to 7 cents in the trumpet–viola test.

There was considerably greater intragroup variability in the tests with tenor voice. For example, the *SD* of individual pitch shift values increased to 14 cents in the tenor voice–viola test and the span of individual pitch shift values increased to 40 cents in the viola–tenor voice test.

The values of DLs (as described above, determined at F0 difference axis as a half distance from 75% 'flat' and 75% 'sharp' response locations) were comparable with the DLs in the tests with the same timbres. The exceptions were again the tests with tenor voice, where the wider range of 'in tune' distributions was expressed also by the bigger DL values (Table 1). The average individual DL of the musicians was 10 cents (*SD* = 4 cents) in the viola–trumpet test but increased to 25 cents (*SD* = 7 cents) in the viola–tenor voice test. The results showed 13 cents (*SD* = 4 cents) in the trumpet–viola test, but increased to 18 cents (*SD* = 4 cents) for the tenor voice–viola test. The values of the group DLs also increased in the tests with tenor voice timbre, reaching 32 cents for the musicians in the viola–tenor voice test (in comparison with 13 cents in the viola–trumpet test), and reaching 52 cents for the non-musicians in the tenor voice–viola test (in comparison with 21 cents in the trumpet–viola test).

Musicians may also take interest in the extent of the F0 deviations, where in at least 75% of cases of the responses from the group, the pitch of the sounds compared would be perceived as 'in tune'. In subtests involving sounds of the same timbre (viola–viola and trumpet–trumpet), the frequency range over which the probability of 'in tune' ratings exceeded 75% only covered a few cents. For the trumpet–viola test (Fig. 4, upper right panel) the maximum number of 'in tune' estimations never exceeded 45%, while the corresponding proportion in the tenor voice–viola test did not exceed 60% (Fig. 5, upper right panel). Thus, when comparing sounds of different timbre, the likelihood that their pitch would be more or less unanimously perceived as 'in tune' decreases, regardless of their F0s.

General discussion

The tests revealed a timbre-induced pitch shift of approximately 15 to 20 cents. Could a deviation of such magnitude be a significant factor for the quality of performance of a musical piece? This value is considerably less than the threshold value above which pitch tends to be associated with the next (or previous) step category of the musical scale, i.e., approximately 50 cents (one quartertone). This conclusion is consistent with intuitive experience from musical practice. On the other hand, we may compare the magnitude of the pitch shift with the difference limen for F0. In the experiment described in this article, the DL of individuals was 8–10 cents for sounds with the same timbre and 10–25 cents for sounds of different timbre for the group of musicians. Thus, the magnitude of the timbre-induced pitch shift was sufficiently large to be perceptible. In fact, several studies have demonstrated that even on F0 difference, a factor of two less than the shift we measured can affect the quality of intonation. For example, Sundberg, Prame, & Iwarsson (1996) studied the intonation quality of vocal performances of Schubert's 'Ave Maria' on numerous commercial recordings and found that the F0 span of the tones should vary by no more than ±7 cents (with a few exceptions) from certain target value (which for some tones still clearly deviated from ETT) in order for all seven highly qualified experts involved in the experiment to perceive the intonation of the performance as faultless. The vocal part was accompanied either by a piano or a harp.

It is conceivable that different individuals perceive a sound differently because of its timbre and may consequently have diverging opinions regarding the pitch of the sound. For example, in addition to what he or she hears reverberating in the room, a singer perceives a considerable part of his or her own voice through the skull (skull/bone conduction). Frequency transmission

in the case of bone conduction differs from that of air conduction (Pörschmann, 2000). Therefore a singer perceives the timbre of his or her own voice differently from other people in the same room.

The aims of the present study did not include investigating the causes of pitch shift. However, for discussion purposes, two possible explanations may be mentioned. First, pitch shift can be predicted by Terhardt's virtual pitch theory using his empirically derived algorithm (Terhardt, 1974; Terhardt, Stoll, & Seewann, 1982a, 1982b). We used Terhardt's algorithm to calculate the virtual pitch of the sounds used in the experiment and estimate their pitch shift. Terhardt's model predicted that the pitch of a trumpet sound would be the equivalent of 16 cents higher than a viola sound with the same F0, and that the pitch of the tenor voice in those conditions would be perceived as 13 cents higher. The predicted values correspond well with the results of our tests, in which a pitch shift of 15 to 20 cents was observed for both pairs.

Second, the described pitch shift in our work tends to support theories by which different aspects of perception, i.e., pitch and timbre, are not independent from each other but exert a mutual influence on each other. The interdependence of aspects of perception, including pitch, timbre, and loudness, has been primarily demonstrated by means of speeded classification tasks (e.g., Melara & Marks, 1990a, 1990b). If one attribute (e.g., timbre) is extracted, it will constitute a 'context' that will influence other attributes (Melara & Marks, 1990a). Russo and Thompson (2005) claim that the influence of such a context is present in the results of perception tests in which the subjective magnitude of a melodic interval constituted by two sounds of different timbre (the amplitudes of lower partials two to eleven increasing monotonically in one and decreasing in the other) was observed to depend on whether or not the timbre difference of the sounds was congruent with the direction of the interval (e.g., dull and bright sounds combined with an ascending step). The conclusions drawn from the present study are consistent with the findings of Russo and Thompson (2005): when the context is a bright timbre (where a sound's spectral center of gravity is located higher), the pitch of the sound is perceived as higher and vice versa – when the context is constituted by a dull timbre (where a sound's spectral center of gravity is located lower), the pitch of the sound is perceived as lower.

Musicians claim that a timbre-induced pitch shift can be anticipated and used to desired effect. For example, according to Kanno (2003), during intonation, violin players deliberately employ slight deviations from standard F0 to create the impression of timbre movement toward brighter or duller.

The fact that a 'brighter' sound is perceived as higher in pitch may be related to another 'top-down' explanation. For example, the timbre of the human voice becomes brighter when the voice is more intense (i.e. as the voice becomes more intense, the amplitude of higher spectral components increases more rapidly than that of lower spectral components – Sundberg, 1987). To increase the level of the voice, one increases air pressure under the vocal cords. This will also automatically increase the F0 of the voice, unless one compensates for that by deliberately using other pitch adjusting mechanisms (Sundberg, 1987). Thus, it appears that, as a result of evolution, we have become accustomed to associating brighter timbre with louder voice, which is the cause of shifts observed in perceived pitch.

Conclusions

The timbre difference of musical instruments is a factor that is likely to influence assessments of the pitch of a sound played on one instrument in relation to the pitch of a sound played on

another instrument, and it is therefore also likely to influence subjective intonation quality in a musical performance. In our tests, the trumpet sound and the tenor voice sound of a professional opera singer, both of which had a bright timbre, were perceived as approximately 15 to 20 cents higher than the viola sound, which had the same F0 but a duller timbre. For the musicians, the F0 difference limen for a pair of sounds consisting of two trumpet sounds or two viola sounds at A3 and with a loudness level of 90 phons was approximately 8–11 cents. The group DL value increased when comparing complex tones with different timbres, and sometimes exceeded 30 cents. The value of the group F0 difference limen for non-musicians was about two times higher than for musicians. Timbre-induced pitch shift occurred among both musicians and non-musicians, although the latter did not distinguish the categories 'in tune', 'flat', and 'sharp' as precisely as musicians.

Acknowledgements

This article has benefited from the support of the Estonian Ministry of Education and Research through targeted financing research grant SF0150004s07. Part of this work was previously presented at the 9th International Conference on Music Perception & Cognition (ICMPC9) in Bologna in August 2006 and at the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM 2009) in Jyväskylä in August 2009.

Notes

1. The database of musical instrument samples is available for public use at: http://theremin.music.uiowa.edu/MIS.html.
2. The algorithm used in this commercial software, as a trade secret, is not overt. The methods for pitch transposition include frequency-domain methods like phase-vocoder, time-domain methods like TDHS (Time Domain Harmonic Scales) or PSOLA (Pitch Synchronized Overlap-add Sample Rate Conversion), and simple sample rate conversion, none of which is free from possible artefacts. The final result depends on the magnitude of the shift and the type of the signal. The frequency-domain methods can cause reverberation, chorusing, or phasiness distortions of the sound. The time-domain algorithms tend to perform poorly with noisy signals, which tend to acquire a buzzy quality (for an overview, see Laroche, 2002). The simple sample rate conversion is free from the aforementioned artefacts; still, the duration of the sound will change and the formant frequencies will shift. Since the amount of pitch transposition necessary for our experiment was extremely small (3% and less) and the signals were monophonic musical tones, the choice of method was not critical. We preferred not to use the 'signal length compensation' and 'formant preservation' commands in the software, as the aforementioned artefacts accompanying those more sophisticated algorithms could cause strangeness of the sound. As the amount of pitch transposition used was very small, the real change in the stimulus length and the shift of the formants of vocal sound were negligible.

References

ANSI. (1994). *American national standard acoustical terminology*. New York: American National Standards Institute.

ASA. (1960). *Acoustical terminology SI, 1–1960.* New York: American Standards Association.

Chuang, C. K., & Wang, W. S. Y. (1978). Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. *The Journal of the Acoustical Society of America*, *64*(4), 1004–1014.

Kanno, M. (2003). Thoughts on how to play in tune: Pitch and intonation. *Contemporary Music Review*, *22*(1–2), 35–52.

Laroche, J. (2002). Time and pitch scale modifications of audio signals. In M. Kahrs & K. Brandenburg (Eds.), *Applications of digital signal processing to audio and acoustics* (pp. 279–310). Boston, MA: Kluwer.

Melara, R. D., & Marks, L. E. (1990a). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics*, *48*, 169–178.

Melara, R. D., & Marks, L. E. (1990b). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 398–414.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego, CA: Academic Press.

Moore, B. C., Glasberg, B. R., & Peters, R. V. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *The Journal of the Acoustical Society of America*, *77*(5), 1853–1860.

Pape, D., & Mooshammer, C. (2006). Is intrinsic pitch language-dependent? Evidence from a cross-linguistic vowel pitch perception experiment. Proceedings of the *ISCA Tutorial and Research Workshop (ITRW) on Multilingual Speech and Language Processing*, Stellenbosch, South Africa.

Pörschmann, C. (2000). Influence of bone conduction and air conduction on one's own voice. *Acustica/Acta Acustica*, *86*(6), 1038–1045.

Russo, F. A., & Thompson W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception and Psychophysics*, *67*(4), 559–568.

Salzberg, R. S. (1980). The effects of visual stimulus and instruction on intonation accuracy of string instrumentalists. *Psychology of Music*, *8*(2), 42–49.

Shonle, J. I., & Horan, K. E. (1980). The pitch of vibrato tones. *The Journal of the Acoustical Society of America*, *67*(1), 246–252.

Singh, P. G., & Hirsh, I. J. (1992). Influence of spectral locus and F0 changes on the pitch and timbre of complex tones. *The Journal of the Acoustical Society of America*, *92*(5), 2650–2661.

Stevens, S. (1935). The relation of pitch to intensity. *The Journal of the Acoustical Society of America*, *6*(3), 150–154.

Sundberg, J. (1978). Effects of the vibrato and the 'singing formant' on pitch. *Journal of Research in Singing*, *5*, 5–17.

Sundberg, J. (1987). *The science of the singing voice*. DeKalb, IL: Northern Illinois University Press.

Sundberg, J. (1991). *The science of musical sounds*. London: Academic Press.

Sundberg, J. (1994). Perceptual aspects of singing. *Journal of Voice*, *8*(2), 106–122.

Sundberg, J., Prame, E., & Iwarsson, J. (1996). Replicability and accuracy of pitch patterns in professional singers. In P. J. Davis & N. H. Fletcher (Eds.), *Vocal fold physiology: Controlling complexity and chaos* (pp. 291–306). San Diego, CA: Singular Publishing Group.

Terhardt, E. (1974). Pitch of pure tones: Its relation to intensity. In E. Zwicker & E. Terhardt (Eds.), *Facts and models in hearing* (pp. 353–360). Berlin/Heidelberg: Springer.

Terhardt, E. (1988). Intonation of tone scales: Psycho-acoustic considerations. *Archives of Acoustics*, *13*, 147–156.

Terhardt, E., Stoll, G., & Seewann, M. (1982a). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *The Journal of the Acoustical Society of America*, *71*(3), 671–678.

Terhardt, E., Stoll, G., & Seewann, M. (1982b). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *The Journal of the Acoustical Society of America*, *71*(3), 679–688.

Vurma, A., & Ross, J. (2007). Timbre-induced pitch deviations of musical sounds. *Journal of Interdisciplinary Music Studies*, *1*, 33–50.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and Psychophysics*, *64*(2), 198–207.

Biographies

Allan Vurma holds a PhD in musicology (2007, with a thesis on 'Voice Quality and Pitch in Singing: Some Aspects of Perception and Production') and is currently a senior researcher at the Estonian Academy of Music and Theatre. His postgraduate degrees are from Tallinn Polytechnic Institute in radio engineering (1978) and from Tallinn State Conservatory (1990) in singing and vocal teaching. Allan Vurma has also been a basso soloist at Estonian Philharmonic Chamber Choir for nearly 30 years.

Marju Raju is a doctoral student of musicology at the Estonian Academy of Music and Theater. She has a MSc in psychology (2007) from Tallinn University and a master's degree in musicology (2008) from the Estonian Academy of Music and Theater. Her main research interests have been pitch perception, singing development, and rhythm and prosody research.

Annika Kuuda has a bachelor's degree in musicology (2007) from the Estonian Academy of Music and Theatre, and she is currently working as a producer of classical radio in Estonian Public Broadcasting.