journal of interdisciplinary music studies
spring 2007, volume 1, issue 1, art. #071102, pp. 33-50

# Timbre-Induced Pitch Deviations of Musical Sounds

*Müzik Seslerinde Tını Kaynaklı Ses Perdesi Sapmaları*

Allan Vurma[1] and Jaan Ross[2]

[1]Estonian Academy of Music and Theatre
[2]University of Tartu and Estonian Academy of Music and Theatre

**Abstract.** This article deals with timbre-induced pitch deviations and their magnitude in environments designed to resemble those that performing musicians encounter in their daily practice. Two experiments were conducted. In the first experiment classically trained singers matched the pitch of synthesized sounds of the piano and oboe. The fundamental frequency of vocal sounds was on average 7 to 13 cents lower than the fundamental frequency of the instrumental sounds. The difference was more pronounced in the case of the piano timbre. In the second experiment, participants compared sounds produced by a single performer from the pitch-matching task of the first experiment to synthesized piano and oboe sounds. A three-alternative forced choice task was used, where participants judged whether the instrumental sounds were higher, lower or equal in pitch when compared to the vocal sounds. Results showed that the highest number of in-tune ratings was elicited if the vocal sounds were performed at about 20 cents below the fundamental frequency of the instrumental sounds. The difference between fundamental frequencies of the sounds perceived as equal in pitch may be explained by different energy distributions in their power spectra.

**Keywords:** Pitch, timbre, intonation, tuning

**Özet.** Bu makale müzisyenlerin günlük çalışma ortamlarına benzer biçimde tasarlanmış ortamlarda tını kaynaklı ses perdesi sapmalarını ve bu sapmaların büyüklüklerini konu alır. İki deney gerçekleştirilmiştir. Birinci deneyde klasik eğitim almış şarkıcılar kendi seslerini sentezlenmiş piyano ve obua seslerinin ses perdeleriyle eşleştirmiştir. Vokal seslerin temel frekansları çalgı seslerinin temel frekanslarından ortalama 7-13 *cent* daha altında gözlenmiştir. Bu fark piano tınısıyla yapılan deneyde daha belirgin çıkmıştır. İkinci deneyde katılımcılar birinci deneydeki tek bir şarkıcının ses perdesi eşleme görevinde ürettiği sesleri sentezlenmiş piano ve obua sesleri ile karşılaştırmıştır. Katılımcıların, çalgı seslerinin vokal seslerine göre daha tiz, daha pes ya da aynı perdede olup olmadığına dair karar verdikleri üç alternatifli sabit bir seçim görevi uygulanmıştır. Sonuçlar vokal sesler çalgı seslerinin temel frekansının yaklaşık 20 *cent* daha altında icra edilmiş olsaydı seslerin aynı perdede olmalarına dair en yüksek oranların elde edilmiş olacağını göstermiştir. Aynı ses perdesindeymiş gibi algılanan seslerin temel frekansları arasındaki fark, bu seslerin güç tayfındaki enerji dağılımlarının farkıyla açıklanabilir.

**Anahtar kelimeler:** Ses perdesi, tını, entonasyon, akort

•*Correspondence:* Jaan Ross, Koidu 122-61, 10139 Tallinn, Estonia;
phone (mobile): +372 522 6886; fax: +372 737 5389; e-mail: jaan.ross@ut.ee

## 1 Introduction

Pitch is one of the most important attributes of musical sounds. In Western music, the majority of musical works are based on one scale or another consisting of steps which during a performance are supposed to be intoned with significant accuracy. Two strategies seem to be available for estimation of the intonation quality in a performance. The first is based on comparison of the sizes of musical intervals to etalon[1] values that are recorded in the long-term memory of the listener. The second is thought to rely on direct comparison of the individual tones assumed to possess the same pitch. This appears to hold equally for perception and production. In practice, sounds are compared to one another already before the beginning of a performance, i.e. in the process of instrument tuning. The reference sound used to tune other instruments is usually given by a tuning fork, the piano or the oboe. Most musicians (except for players of instruments with so-called fixed tuning, such as the piano) need to set intonation standards not only before the beginning of a performance, but also during it.

The meaning of the word ‘pitch’ is fraught with a certain degree of ambiguity. Quite often, such as in descriptions of musical scales, the pitch of a sound is considered equivalent to its fundamental frequency. In a harmonically complex tone, the waveform period length is inversely proportional to its fundamental frequency, and the partial frequencies are integer multiples of the fundamental. For this reason, two simultaneous harmonic complex tones can be tuned to sound in unison relatively easily. The tuner simply needs to minimize the number of audible beats which are amplitude variations due to interference between the two waves. The frequency of the beats corresponds to the difference between the fundamental frequencies of the two waves (a complete lack of beats shows perfect unison). This method, however, does not work when the sounds are performed with vibrato, which makes it almost impossible to discern beats.

Should one, however, choose to treat the category of pitch as primarily perceptual, its link to the fundamental frequency becomes less pronounced. According to the definition of the American Standards Association, pitch is that attribute of sounds that permits the organization of sounds into a musical scale (ASA 1960). Another definition of the American National Standards Institute states that pitch is a characteristic of perceived sounds that makes it possible to order the sounds on a scale from low to high (ANSI 1994). Human perception is always to a smaller or greater extent subjective—it depends on the perceiving individual and the context in which the act of perception takes place. Although the perceived pitch of a sound is to a large extent determined by the fundamental frequency involved, it is also influenced by other features of the sound, such as timbre and/or pressure level (Terhardt 1988). It is therefore possible for two sounds having identical fundamental frequency values but different timbres to be perceived by listeners as having different pitches. It is likewise possible that two sounds that are judged equal in pitch when presented consecutively will not be perceived as such when presented simultaneously.

---

[1] The standards of intonation for a given culture are the learned interval categories of the scales of that culture (Burns 1999).

Terhardt (2000) calls changes in pitch resulting from the above factors “pitch deviations.” The magnitude of such deviations, however, is not very large. In Terhardt’s (1988) estimate it does not usually exceed approximately 50 cents,[2] or one-quarter of a tone. From the point of view of Western music performance, pitch deviation by a quarter tone must be regarded as far from insignificant. Experimenters have measured sine tone pitch difference limens of ten times less, i.e. about five cents at the frequency of 250 Hz (e.g., Moore 2003: 198).

According to the definition of the ASA (1960), the timbre of a sound is that attribute of auditory sensation which allows a listener to differentiate between two sounds that are presented in a similar manner and have the same loudness and pitch. The difference in timbre allows us to distinguish between different musical instruments performing the same note. Traditionally, timbre is linked to the distribution of energy in the power spectrum. For example, the property of timbre called brightness is correlated with the location of the energy centroid on the frequency axis (Risset and Wessel 1999: 147). Still, there are many other features of a sound that also contribute to the perception of timbre, such as the temporal patterning of the parameters of the sound and the presence of different noise components (Moore 2003: 270).

This article is the outcome of an investigation into timbre-induced pitch deviations and their magnitude in environments designed to resemble those that performing musicians encounter in their daily practice. The environments in question are expected to vary widely. For example, a singer preparing for a solo in a musical work of complex texture may choose to track a single instrument in the orchestra in order to use it as a reference for establishing the pitch of the initial note in the part (s)he needs to start with. If timbre has a sufficiently strong influence on perceived pitch and if listeners elect to concentrate on the unison between the singer’s voice and an instrument characterized by a timbre which is different from that of the instrument selected by the singer as reference, the listeners may perceive the singer’s pitch as off. A listener may also concentrate on detecting the beats between two simultaneous sounds that are supposed to be in unison, or may instead try to compare the pitches of successive sounds that are represented by the same musical note in the score but sung or played in different timbres. These strategies may lead to different conclusions about the accuracy of the intonation of the performance.

In previous studies, the stimuli used in the experimental investigation of timbre-induced pitch deviations have mostly involved sounds quite unlike any of those produced by real musical instruments or the human voice. In a number of cases, timbre-induced pitch deviations have been studied by using indirect methods, and not by comparing musical sounds as they occur in real performances. Thus, Singh and Hirsh (1992) used stimuli consisting of harmonic partials whose amplitudes were set larger than zero only in one particular frequency region. They found that shifting the locus of that region on the frequency axis had a significant effect on the perceived pitch of the sound. They also observed that the perceived pitch of a sound was reported to rise when this locus was moved toward higher frequencies, even when the fundamental frequency of the stimulus was decreased at the same time. This phenomenon occurred

[2] Cents are logarithmic units for measuring pitch and fundamental frequency. 100 cents are equal to one semitone.

only when the frequency changes remained within the range of two to four per cent, however.

Warrier and Zatorre (2002) used harmonic complex tones consisting of 11 partials. The tones were modified in two different ways: first by making the partials increase or decrease monotonically in amplitude, and secondly by increasing the amplitudes of partials one to six while reducing the amplitudes of higher partials. The fundamental frequencies were increased by 17, 35 and 52 cents compared to that of the reference tone. As a result, complex tones with different spectra but the same fundamental frequencies were perceived as having different pitches. The perceived difference was smaller when the tones were presented as part of a melody.

Russo and Thompson (2005) found that a melodic interval was perceived as wider when its higher component was brighter than the lower one, and vice versa. They argued that this phenomenon can be explained as an illusion arising from the interaction between the pitch and the timbre of the component sounds of the interval and that timbre creates a supplementary context which influences the extraction of pitch information. Worthy (2000) found that wind instrument players trying to match a given pitch, tended to produce a tone with a slightly higher fundamental frequency than the pilot tone when the latter had brighter timbre, and with a slightly lower frequency when the pilot tone was duller. In the experiments of Ogawa and Murao (2004), music students were asked to use their voice (in both the modal and falsetto registers) to reproduce musical intervals consisting of sine tones, piano sounds or female vocal sounds without vibrato. The results obtained were dependent of both the timbre of the pilot sounds and the vocal register used in the reproduction task. Platt and Racine (1985) investigated the ability of informants to tune their instruments, i.e. to set the pitch of a sound to match that of another reference sound. They concluded that this ability deteriorated when the sounds involved possessed different timbres.

Terhardt (1971) developed the so-called pattern matching theory that is aimed at modeling human pitch perception on the level of the central auditory system. The basic ideas underlying his model are, first, that the perceived pitch of a sound results from matching its real spectrum to an internally stored pattern of a harmonically complex sound and, second, that the internal pattern of a complex sound has somewhat spread out partials because of the interaction between them. According to his theory (as well as the results of the experiments he conducted), the pitch of a complex sound with strong higher harmonics is shifted somewhat downwards in comparison with the pitch of a pure tone corresponding to the same fundamental frequency. However, the results of later research by Hartmann and Doty (1996), and Peters et al. (1983) did not confirm these findings.

To obtain the data for the present study, we conducted two experiments investigating pitch deviation by comparing pairs of sounds with different timbres in quasi-realistic environments that closely resemble real situations occurring in musical practice. In the first experiment, participant singers were asked to match the pitch of computer-synthesized piano and oboe sounds. In a task like this, singers must compare the pilot pitch with the pitch of their own voice in order to produce the best match. The second experiment involved a comparison task in which participant listeners assessed the pitch of sounds having the same or close fundamental frequency values but different timbres. The triple of timbres were used here as in the first experiment. Listeners judged whether the vocal sounds were higher, lower or equal in pitch as compared to the instrumental sounds.

## 2 First Experiment: Method, Stimuli and Participants

Classically trained singers were asked to use their voices to match pilot sounds with piano and oboe timbres. This test was aimed at investigating, firstly, whether the timbre of the pilot sound influences the outcome of the task and, secondly, the matching accuracy that singers can achieve. It was hypothesized that, due to the brighter timbre of oboes, a sound with the timbre of an oboe would be perceived as higher in comparison to a sound with the timbre of a piano having the same fundamental frequency. It was further hypothesized that, as a consequence, the oboe sounds would be matched to pitches somewhat higher than those of the piano sounds. The experiment was performed with seven participants: two tenors, three basses, one soprano, and one contra-alto. All had University level musical education. Five are employed in a highly reputed professional chamber choir and two are freelance artists mostly working with international projects. The participants' average age was 37 (the youngest was 23 and the oldest 49 years old, as noted in Table 1).

**Table 1.** Average data on singers' matches of the pilot tone (synthesized piano or oboe). Columns 4 and 6 show the mean difference between the fundamental frequencies of the pilot and the singer's match, together with standard deviation (SD; columns 5 and 7). Δ in column 8 shows the difference between the average matches to the piano and oboe sounds in columns 4 and 6, respectively. All numbers in columns 4 to 8 are in cents. When the average difference of the match from the pilot sound is statistically insignificant, according to a one-sample t-test, its value is presented in italics. Data for the participant MT, who in his performance exhibited the largest deviations from the pilot sounds, is presented in bold.

| Singer | Age (years) | Voice category | Piano | | Oboe | | Δ |
|---|---|---|---|---|---|---|---|
| | | | Mean | SD | Mean | SD | |
| MK | 23 | sop | *1* | 17 | 9 | 15 | 8 |
| KS | 33 | c-alto | *2* | 28 | *-6* | 26 | -8 |
| **MT** | **36** | **ten** | **-37** | **14** | **-39** | **15** | **-2** |
| ML | 35 | ten | -6 | 18 | *5* | 27 | 11 |
| KK | 49 | bass | -18 | 18 | *-6* | 22 | 12 |
| RL | 40 | bass | -28 | 37 | -19 | 37 | 9 |
| UJ | 41 | bass | *-4* | 20 | *5* | 32 | 9 |
| Mean | 37 | | -13 | 22 | -7 | 25 | 6 |

The pilot sounds were generated by Microsoft MIDI Mapper, with a sampling frequency of 22,050 Hz. We preferred computer synthesis of pilot sounds to natural production because the procedure is significantly simpler, the parameters of the sounds are set by the investigators and it is possible to produce sounds with pitches outside the natural range of the instruments. The piano and the oboe were chosen as the pilot timbres because singers encounter these instruments very frequently in their everyday performance practice. The piano is the most common accompaniment instrument for singers. The oboe, in turn, has a uniquely penetrating timbre, which gives it the ability to be audible over other instruments in large ensembles and makes it easily heard for tuning.

The spectral characteristics of the piano and oboe sounds differ. In the piano spectrum, the highest amplitude usually belongs to the fundamental, while the next five harmonics demonstrate monotonically decreasing intensities (see Figure 1). In the oboe spectrum, the intensity of harmonics increases slightly from component one to four, the latter being about 10 dB louder than the fundamental. Perceptual experiments have indicated that the oboe timbre is perceived as brighter than the piano timbre (Krumhansl 1989). This is explained by the fact that the middle harmonics of the oboe sound more intense than the lower ones.

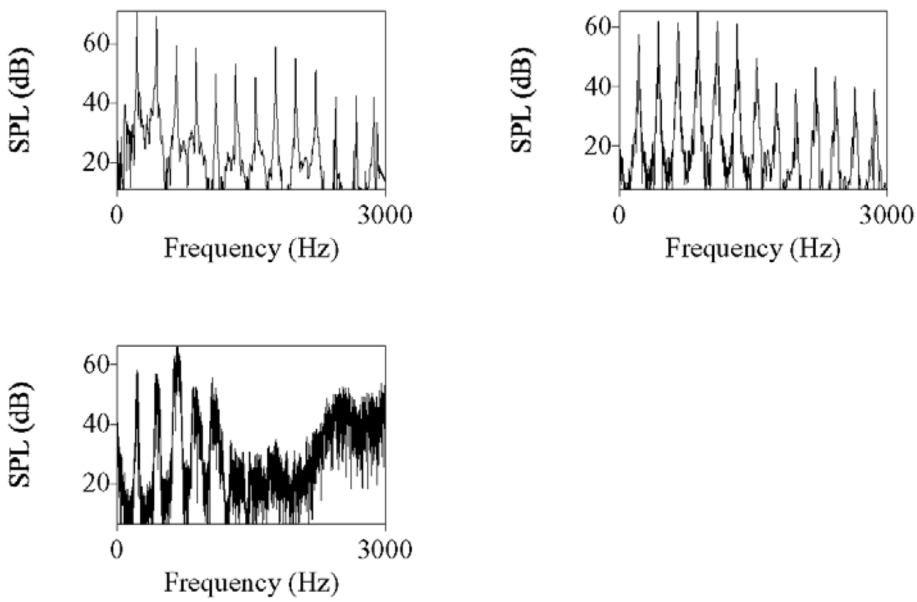


**Figure 1.** FFT spectra of sounds used in experiments. Top left: piano; top right: oboe; bottom: singing voice.

The duration of the quasi-stationary parts of the synthesized sounds used in our experiment was approximately 100 ms, which roughly corresponds to the duration of an eighth note performed in allegro tempo. The synthesis algorithm treated the piano and the oboe sounds differently. The fundamental frequency of the piano sound dropped by approximately 10 cents during the offset part of the sound, which had a duration of about 400 ms. The fundamental frequency of the oboe sound wavered by ± 5 cents during the attack portion of the sound (approximately 20 ms), and stabilized for the stationary part (see Figure 2, bold curves top left and right). The fundamental frequencies of the sounds were measured as averages over the samples covering their quasi-stationary parts. Fundamental frequency curves were obtained using the speech analysis freeware Praat4 (www.praat.org). Frequency calculations were based on the autocorrelation method.

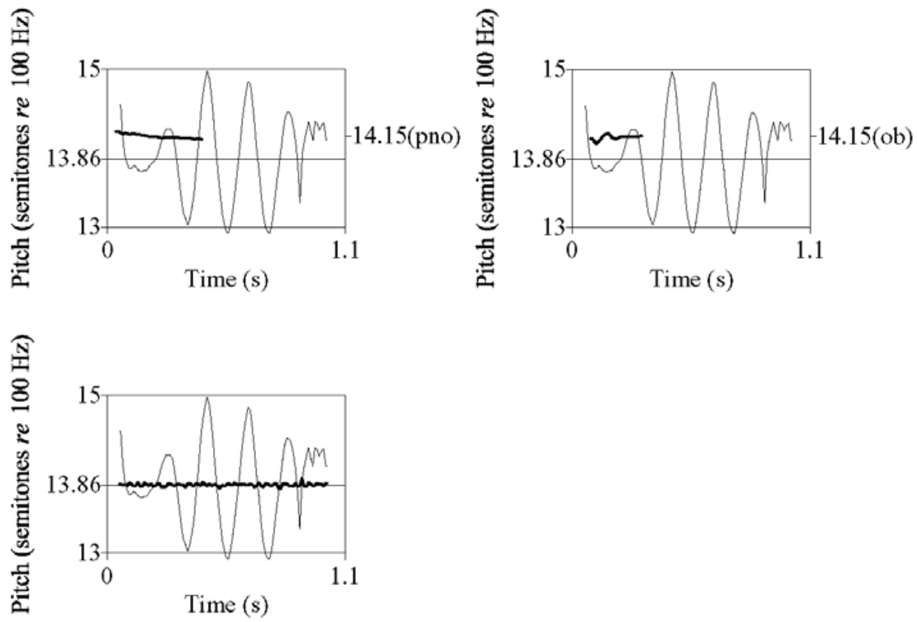


**Figure 2.** The fundamental frequency contours of sound pairs. Top left: piano (bold) and singing voice, sound A; top right: oboe (bold) and singing voice, sound A. Both panels represent the best match between the vocal and instrumental sounds in the second experiment, so that F0 of the former is 13.86 semitones and F0 of the latter 14.15 semitones higher the 100 Hz reference level. Bottom: singing voice sound A and its modification A' without vibrato (bold).

Participants were asked to match the pitch of a pilot sound with their voice by singing the vowel /a/. The time for performing the task was not restricted. Nonetheless, all singers fulfilled it with no hesitation, and began performing within the second immediately following the presentation of the pilot, as soon as they had filled their lungs with air.

The following fundamental frequencies were used in the experiment: D2 (73.4 Hz), D3 (146.8 Hz), A3 (220 Hz), D4 (293.7 Hz), A4 (440 Hz), D5 (587.3 Hz) and F5 (698.5 Hz).[3] Some of these are outside the range that can be produced on a natural

[3] The values were calculated according to equal temperament.

oboe. In addition to sounds with the fundamental frequencies indicated, additional sounds were synthesized with F0 values 25 and 50 cents higher and lower than the base fundamental. This was accomplished by means of the pitch correction subroutine of the WaveLab 4.0 (© Steinberg Media Technologies GmbH) sound processing software. The resulting five variants for each fundamental frequency value were performed both with the piano and the oboe timbres, and repeated three times during the same testing session. The order of the stimuli was randomized. Each singer was presented with four of the seven fundamental frequencies listed above, together with their synthesized modifications. As a result, each singer had to perform pitch matches in a total of eight successive testing sessions (four with oboe and four with piano timbres), each consisting of 15 sounds. The four fundamental frequency values were chosen so as to cover the range of vocal production of each singer as fully as possible. A laptop computer equipped with a Realtek AC97 sound card and Sennheiser HD590 open design earphones was used in the experiments. Open design earphones were chosen in order to make it possible for singers to be able to hear their own voices when singing. The participants' performance in the matching task was recorded with a sampling frequency of 22,050 Hz using the SONY TCD10 DAT-recorder and an AKG420 head microphone. The distance between the microphone and the corner of the singer's mouth was 3 cm. The experiments were conducted at the Tallinn Philharmonic Society facility, in a rehearsal room characterized by short reverberation time. In most cases, singers matched pilot sounds by singing for approximately one second. The fundamental frequencies of matching sounds were measured during the stable part of each sound.

## Results

The results of the first experiment are presented in Table 1. Responses by individual singers did not form a homogeneous group. Apart from two cases (singers MT and RL), the differences between the fundamental frequency values of the pilot tones and the participants' matches were relatively small. Standard deviations seem to correlate poorly with singers' average deviation from the fundamental frequency of the pilot tone. For example, they remain between 14 and 17 cents for the singers MK and MT, who otherwise exhibit differences in their accuracies in matching the pilot tone. For MK, the average values of matching the piano and oboe tones were 1 and 9 cents respectively, while for MT, the same values were minus 37 and minus 39 cents.

In order to test the statistical significance of differences between the fundamental frequencies of the pilot and its match, a one-sample t-test was performed. These differences were not significant (at 95 percent confidence level) in 7 cases out of the total 14 (see Table 1 the numbers in italics in columns 4 and 6). The values of those cases remained between minus 6 and plus 5 cents. The rest of the differences between the pilot and the matching tones were negative in 6 cases (i.e. the fundamental frequency of the match was lower than that of the pilot), ranging between minus 18 and minus 39 cents, and positive in one case (singer MK's match to the oboe timbre, with a difference of 9 cents). It can be concluded that, in Experiment I, singers matched the fundamental frequencies of the piano and oboe pilot sounds at frequency values equal to or somewhat lower than those of the pilots. The tendency to produce a fundamental frequency lower than that of the pilot was more pronounced with the piano-timbre sounds. The aggregated average values for the 7 participating singers amount to minus 13 (piano timbre) and minus 7 (oboe timbre) cents. This difference is statistically significant according to the Mann-Whitney rank sum test ($p = .003$).

## 3 Second Experiment: Method, Stimuli and Participants

The results of the first experiment demonstrated that timbre differences between sounds sharing the same fundamental frequency tend to cause statistically significant differences in vocal matching of these sounds by professional singers. At the same time, inter-individual differences between the singers who participated in the pitch-matching task of the first experiment were sometimes greater than the deviation attributable to timbre variation (see Table 1). This suggested that it is also necessary to study differences in intonation between individual singers in a more systematic way.

In Table 1, the results of the singer MT, a tenor, may be singled out because they exhibit the largest deviations from the pilot sounds' fundamental frequencies (minus 37 for the piano sounds and minus 39 for the oboe sounds). We hesitate to attribute this to a simple lack of ability to sing with correct intonation on the part of MT because of his excellent professional reputation. MT is known for his perfect solfège skills, long performance practice and brilliant vocal timbre. In this context, it would be reasonable to suppose that timbre exerts an influence on accuracy in matching a given fundamental frequency, and that differences between the intonation accuracies of individual singers may be caused by the timbre differences between their voices. When singers perform the task of matching the pilot tone with their voice, the feedback circuit regulating the pitch of production involves comparing the pitches of the two tones (the pilot tone and the singer's own voice). This is supported by the findings of Burnett et al. (1997), which note that the auditory channel retains a leading role in spite of the fact that feedback through kinaesthetic sensations is also important. A close connection between perception and production of sounds is therefore expected to occur during the pitch matching task, and mismatches may be at least partly, attributed to perceptual mechanisms. It could be hypothesized that, due to the difference between the pilot and the matching sound timbres, the best perceptual match between the pilot sounds and the matches sung by MT occurred when the latter were 35 to 40 cents lower than the respective pilots.

In order to test this hypothesis, a second experiment was designed and performed. Four randomly selected sounds performed by MT himself in the first experiment were used as the stimuli. We will refer to these as 'the vocal sounds,' and designate them as A, B, C, and D. The fundamental frequencies of the vocal sounds were located in the vicinity of A3 (see the description of Experiment I above), which lies approximately at the middle region of MT's voice range. The average values of the vocal sounds measured over the stationary part of the tone, including the full vibrato periods, were 222.8, 224.9, 214.6, and 217.7 Hz respectively. The duration of the sounds was about one second and the amplitude of the vibrato about one semitone, or 6 percent of F0. The frequency of the vibrato was approximately 5.6 Hz (see also Figure 2).[4]

In this experiment, participants were asked to compare the vocal sounds produced by MT with a synthesized sound having the timbre of a musical instrument (the piano or the oboe), and to rate the vocal sound as equal in pitch, sharp or flat in comparison to the reference synthesized sound. The sounds synthesized were analogous to the pilot sounds in the first experiment. They included the core sound A3, with a fundamental frequency of 220 Hz, as well as its F0 modifications extending by 25 cent steps to plus and minus 100 cents, amounting to a total of 9 comparison stimuli differing slightly in their fundamental frequencies. However, not all of these were actually used in the experiment. In order to optimize the duration of experimental sessions, the nearest synthesized pitch match to each vocal sound was determined before the experiment, and was then narrowed to include only that match and the two steps immediately below and above it. In this way, a range of 100 cents was covered in comparing the pitch of the vocal sound with the pitch of the synthesized sounds. The nearest match was determined by a pilot experiment involving a single participant.

The experiment was conducted by using the perception experiment module in the well-known Praat4 software. Listeners had to record their rating of the correspondence between the reference sound and the stimulus by clicking one of the three buttons (labelled as flat, sharp, or in tune) displayed on the computer screen. The response time was not limited. There was a pause of unlimited duration allowed after every 10 comparisons. The order in which the sound pairs were presented was randomized. Every pair occurred twice during each session and identical pairs were not allowed to occur contiguously. The two sounds forming a pair were separated by a silence of 2.5 seconds.

A singer is never able to maintain a perfectly constant fundamental frequency or vibrato during the whole sound production cycle because the pitch of his or her voice is influenced by instabilities in muscle innervation, as well as fluctuations in cardiovascular and lymphatic activity (Titze, 1994). For this reason, we repeated the test of the present experiment using the same piano-timbre sounds, but artificially modified the vocal sounds A, B, C, and D in order to free them from vibrato. This was accomplished by means of the 'Stylize Pitch' subroutine of Praat4. The modified sounds A', B', C', and D' had no vibrato, and their fundamental frequencies remained constant within the limits of ±4 cents (see Figure 2, bottom: bold curve). The spectral properties of the modified sounds were similar to those of the original sounds. They contained an amplitude vibrato of about 1 dB. The modified sounds, however, created a somewhat artificial impression on the experimenters.

There were 20 participants, 11 male and 9 female. Their ages ranged from 20 to 49 years, the average being 36 years. All were professional musicians, including students of the Estonian Academy of Music and Theatre, choir singers, conductors, ear training teachers, and a sound engineer. None of them reported the ability to perceive absolute pitch, nor did they report any hearing disorders.

[4] Values of up to 6 or even 12 percent for the amplitude of the vibrato and between 5 and 8 Hz for its frequency are typical of Western classical voices (Sundberg 1994).

## Results

The participants' ratings of the match between the vocal sounds and the synthesized piano-timbre sounds are presented separately for the four vocal sounds A, B, C, and D in Figure 3. The peak of 'in tune' responses corresponds to the F0 value of the vocal sound produced by MT at about 20 cents less than the F0 value of the piano-timbre sound. The number of 'flat' and 'sharp' ratings is nearly equal at that point. The number of 'sharp' ratings begins to increase with lower F0 values of the piano-timbre sound and, conversely, the number of 'flat' answers starts to increase with the F0 value of the piano sound rising. The number of 'in tune' answers amounts to 80 to 90 percent of the total at the best matching point (F0 difference of 20 cents below the piano-timbre sounds).

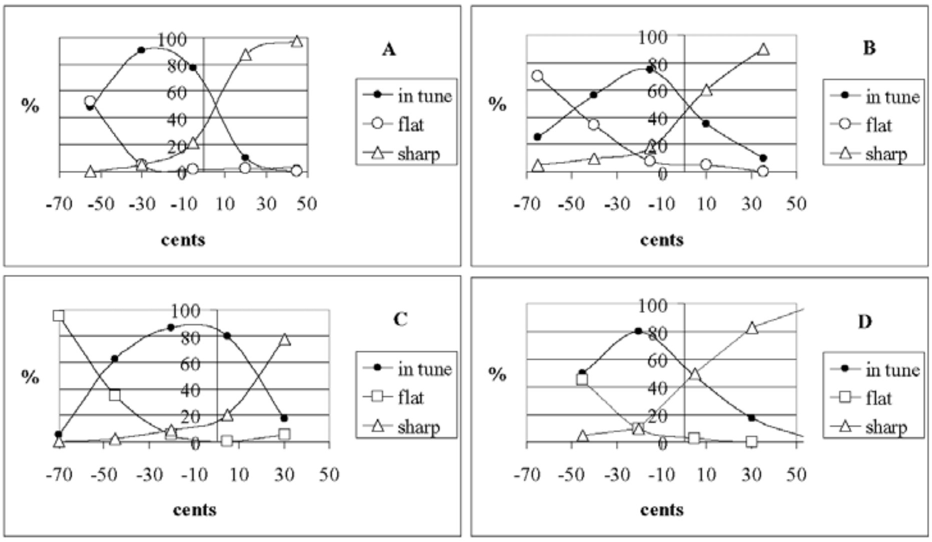


**Figure 3.** Distribution curves of 'in tune', 'flat' and 'sharp' ratings after comparison of the vocal sounds to the synthesized piano sounds. Horizontal axis: fundamental frequency of the vocal sound minus fundamental frequency of the instrumental sound; vertical axis: percentage of 'in tune', 'flat' or 'sharp' ratings. Top left: vocal sound A, top right: B, bottom left: C, bottom right: D.

We further tried to establish the threshold values beyond which the two sounds with different timbres (piano and voice) are no longer perceived as having equal pitch. The thresholds were set at 75 percent of the consensus level (i.e. at the level where at least 75 percent of listeners rate the sounds as differing in pitch). The upper limit of the subjective equality zone is located at F0 plus 14 to 27 cents (i.e. the vocal sound is higher than the piano-timbre sound by 14 to 27 cents), and the lower limit is located at minus 59 to 67 cents (i.e. the vocal sound is 59 to 67 cents lower than the piano sound). This means that, given intonation differences spanning a region as wide as approximately 80 to 90 cents (the subjective equality zone), a vocal sound would still be perceived as equal to a piano-timbre sound by at least 25 percent of listeners.

Terhardt (1988) has distinguished between the sensory and harmonic purity of intervals, including the unison. According to him, an interval is characterized by sensory purity when it is not disturbed by beats or roughness. On the other hand, a melodic or a harmonic interval is characterized by harmonic purity when its size coincides with its memorized pitch-interval templates which in turn are partly of natural, partly of cultural origin. In many cases, an interval cannot be pure according to both sensory and the harmonic criteria at the same time. A performer needs to strike a compromise in such instances, which implies that the so-called correct intonation is a multidimensional category.

The comparison of vocal sounds with the oboe-timbre sounds yielded results that were substantially similar to those described above for the piano sounds. Figure 4 compares the distributions of 'in tune' answers both for the oboe and the piano sounds. The oboe-timbre results show a pitch deviation going in the same direction and having about the same magnitude as those obtained for the piano-timbre sounds. The distribution curves in Figure 4 indicate that the magnitude of the pitch shift for the oboe sounds is somewhat less than that for the piano sounds. The thresholds of the subjective equality zone for the oboe sounds (see analogous explanation for the piano sounds in the paragraph below Fig. 3) remain between plus 20 to plus 35 cents, and minus 53 to minus 69 cents, covering a region of 73 to 99 cents.

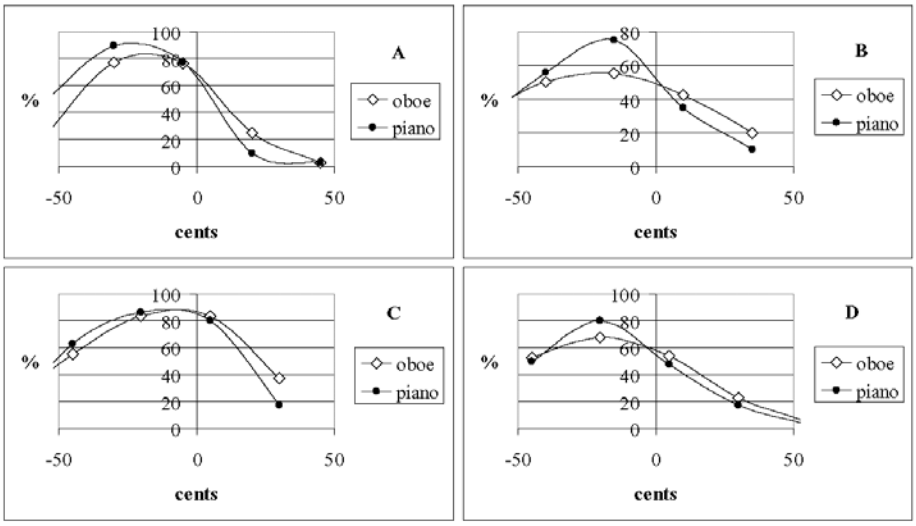


**Figure 4.** Distribution curves of 'in tune' ratings after comparison of the vocal sounds to the synthesized piano and oboe sounds. Horizontal axis: fundamental frequency of the vocal sound minus fundamental frequency of the instrumental sound; vertical axis: percentage of 'in tune' ratings. Top left: vocal sound A, top right: B, bottom left: C, bottom right: D.

Comparing the results obtained in the first and the second series of experiments, we see that there is a discrepancy in the magnitude of the pitch shift between the vocal sounds produced by MT and the synthesized sounds. In the first experiment, the F0 of the tones matched by MT were 35 to 40 cents lower than the F0 of the pilot tones. Yet in the second experiment, the differences were nearly two times less, i.e. approximately 20 cents. It may be proposed that a singer perceives her/his own voice differently from other listeners because, in addition to the normal auditory pathway, (s)he receives the sound also by way of bone conduction. Experiments by Pörchmann (2000) indicate that the parts of the sound spectrum falling below 0.7 kHz and above 1.5 kHz may be more accessible to the listeners than to the singer her/himself. We cannot exclude the possibility that the singer may have perceived the pitch shift as larger than the other listeners did. There is, however, no data that directly permits an explanation of differences in the magnitude of pitch shifts on the basis that the perceptual experience of a vocal sound differs between its performer and other listeners.

It is worth noticing that in three cases (A, B and D) out of four in Figure 4, the peak of the 'in tune' responses curve for the piano timbre, compared to the oboe timbre, was 10 to 20 per cent higher. This difference may be caused by the fact that piano sounds are a habitual presence in the daily practice routine of performing musicians, which is to say that they encounter piano sounds much more frequently than oboe sounds. Another factor contributing to this result may have been the slightly shorter duration of the oboe sounds compared to the piano sounds, arising from the fact that offsets for the former were synthesized as steeper.

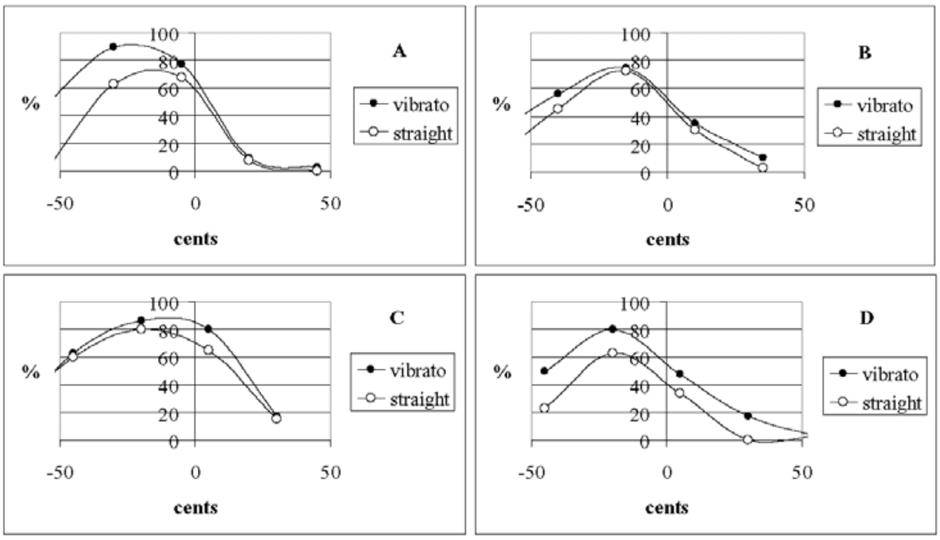


**Figure 5.** Distribution curves of 'in tune' ratings after comparison of the vocal sounds with and without vibrato to the synthesized piano sounds. Horizontal axis: fundamental frequency of the vocal sound minus fundamental frequency of the instrumental sound; vertical axis: percentage of 'in tune' ratings. Top left: vocal sound A, top right: B, bottom left: C, bottom right: D.

Figure 5 presents a comparison of 'in tune' answers from the two different versions of second experiment, in which piano-timbre sounds were compared to vocal sounds with and without vibrato. The graphs for all four sounds (A, B, C, and D, and their modifications A', B', C', and D') have a similar shape, exhibiting a pitch shift of about 20 cents' magnitude. This shows that vibrato has little or no effect on pitch deviation between vocal sounds and synthesized piano sounds that have the same fundamental frequency.

## 4 General Discussion

In this paper, we have described two experiments aimed at studying pitch deviation that can be demonstrated to occur in comparisons of two sounds having identical or close fundamental frequencies, but different timbres. The stimuli in the experiments included (1) digitally synthesized sounds designed to resemble natural instrument sounds (‘quasi-natural’ sounds), (2) sounds produced by the human singing voice, and (3) digitally manipulated vocal sounds. The participants of the pitch matching as well as accuracy rating tests were professional musicians.

The results of experiments where the pitches of sounds with different timbre were compared to each other at the region of 220 Hz (the frequency of A3) show relatively similar magnitudes (approximately 20 cents) of pitch shifts. The explanation for the pitch shift seems to be caused by different energy distributions in the sound spectrum. The sounds which transmit more energy at higher frequencies (and whose timbres consequently comes across as brighter) are perceived as having pitches higher than those of sounds which convey more energy at lower frequencies (whose timbres come across as duller). In other words, to the end that sound with bright timbre would be perceived as equal in pitch with successively presented sound with dull timbre, the fundamental frequency of the former should be approximately 20 cents lower.

This explanation is at odds with Terhardt’s (1971) theory and data. His theory predicts that the pitch of a sound with stronger high harmonics would be perceived as somewhat lower compared to the pitch of a sound with the same fundamental frequency but weaker upper harmonics. The results of the experiments described above, however, are in agreement with other similar studies reviewed in the introductory section of this paper. Those studies tend to link timbre-dependent upward pitch shifts to the upward movements of the spectral center-of-gravity.

Natural sounds, as produced by musical instruments or the human voice, are never completely stable, even within individual notes, particularly during the attack and decay portions. This is because a performer is technically unable to guarantee a note’s stability due to the complex nature of tone production (Risset and Wessel 1999). However, it is common to attribute a single pitch label to a musical note, even when its fundamental frequency is far from stable. This makes it possible to compare different notes in terms of their relative position on the pitch scale, although experienced musicians may, if they so choose, be able to perceive the so-called micro-fluctuations of fundamental frequency.

In the present study, we have attributed pitch values to sounds on the basis of the average fundamental frequency measured over the quasi-stationary part of the F0 curve of these sounds. We cannot exclude the possibility that, in certain situations, pitch may be attributed to sounds according to a different mechanism that is not necessarily based on average fundamental frequency calculations. Pitch shifts of the magnitude of around 20 cents obtained in this study, however, do as a rule exceed by several times the fundamental frequency instabilities that can be detected within individual notes as random or quasi-random fluctuations. Vocal sounds performed with vibrato represent an exception to this rule. Nevertheless, many earlier studies (e.g., Sundberg 1978; Shonle and Horan 1980) appear to have reached the consensus that the pitch of a tone with vibrato is perceived at the average fundamental frequency. This suggests that the pitch shifts observed in this study should not be treated as artefacts due to a possible alternative relationship between fundamental frequency and pitch in real musical performance, even though the precise magnitude of the shifts observed may require some correction in the future.

The results obtained must not be automatically generalized to apply to other domains of musical sound. The authors of the present article have only examined a restricted fundamental frequency region, and have further limited their investigations by using sounds produced by a restricted number of musical instruments. Also, the pitch comparison of investigated timbres and sine tone would be desirable in the future.

An important issue to consider relating to the observed pitch shifts is their relevance to musical performance practice. It is likely that this depends on the particular situation and on the cognitive attitude of the listener. At the same time, an order of magnitude of about 20 cents, as established in this study, must be regarded as rather large, representing four times the frequency difference limen of about 5 cents. Nevertheless, it may not stand out, since it is well known that listeners tend to perceive musical intervals in a categorical manner, which means that human auditory perception is less sensitive to pitch differences falling within the category of the reference sound. The limits of a perceptual category are usually much wider than 20 cents (Burns and Ward 1978). It is also significant that the accuracy of trained singers in producing a specific pitch is usually no better than about 20 cents (Mürbe et al. 2004), and that the accuracy of violinists is no better than 10 cents (Brown and Vaughn 1996). Since the perception of pitch depends on timbre, performing musicians are not expected to adhere strictly to the equally-tempered fundamental frequency values in tuning their instrument or voice (e.g. by an electronic tuning aid). Apparently, the quality of perceived intonation can sometimes be improved by changing the timbre of the sound instead of its fundamental frequency. Singers, for example, may choose to manipulate the quality of their vowels in both the front/back and the open/close dimensions as well as by varying the level of the singer's formant[5]. Of course, the impact of timbre on pitch perception remains rather limited.

During the matching of pilot sounds with voice in the first experiment, deviations from pilots varied depending on the timbre of the pilot sound (piano versus oboe). However, the deviations between the matches were of much smaller magnitude than those apparent in perception, and can be assimilated to just noticeable differences in the frequency domain. We assume that a singer may freely choose an instrument to track for intonational reference during the performance. The results obtained might have been different for sounds of other timbres. Also, additional information in this regard can be obtained in the future by comparing the sounds produced by acoustic

---

[5] In phonetics the quality of a vowel depends on the position of the tongue in the mouth cavity (if the tongue is in the front of the mouth, then the frequency of the second formant in the vowel spectrum will be higher) and on the degree of mouth opening (if the mouth is more open, then the frequency of the first formant will be higher). The presence of the singer's formant is typical to classically trained voices and it gives brightness and carrying power to the voice.

pianos and oboes. Although results from the second experiment demonstrated that the voice of the tenor MT matched the oboe and piano timbre tones best when the singer's voice was 15-20 cents lower than the F0 of the oboe and piano tones, it does not necessarily imply that the same would hold for the voices of other singers. Also, the hypothesis that the timbre of singer's own voice can have an influence on his or her intonation requires further testing.

The main cause of the phenomenon of pitch deviation in present experiments seems to be the difference in spectral envelope of the compared sounds. In psychoacoustic terms, this means that the location of the center of gravity of a sound spectrum on the frequency axis may lead to the pitch of that sound being perceived as higher or lower than its actual F0 value. In principle, it would be possible to calculate the location of the center-of-gravity in a spectrum and, in this way, to attempt to produce a computational model of pitch shift dependence on the spectral center-of-gravity. The sensitivity of the human auditory system, however, varies at different frequencies, and may further vary in considerable degree from individual to individual. For this reason, we have refrained from attempting to construct such a model within the limits of the present article.

## 5 Conclusions

Two sounds with different timbres but identical or close fundamental frequency values may be perceived by listeners as having different pitches. Results of the first experiment demonstrated that professional singers matched the fundamental frequencies of the pilot tones at F0 values that were equal to or slightly lower than those of the pilots. This tendency was somewhat more pronounced in matching the piano-timbre sounds than in matching the oboe-timbre sounds. Results of the second experiment demonstrated that, in the rating tests, the F0 difference at the subjectively equal pitch of the vocal and instrumental sounds corresponds to approximately 20 cents. In other words, in order to be perceived as equal in pitch to the vocal sounds, the instrumental sounds must have a slightly higher fundamental frequency. Elimination of vibrato from the vocal sounds did not have a significant effect on the results. The observed pitch deviations may be due to different energy distributions in the power spectra of sounds, which may result in different locations of the spectral center-of-gravity. The latter is in turn thought to cause small deviations in the perceived pitch of the sounds of different timbre.

## Acknowledgements

This research has been supported by the Estonian Science Foundation grant no. 4712. The authors gratefully acknowledge the assistance of Meelis Leesik in the language editing of this paper.

## 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.

Brown, J. C. and Vaughn, K. V. 1996. Pitch center of string instrument vibrato tones. *Journal of the Acoustical Society of America*, 100(3): 1728-1735.

Burnett, T. A., Senner, J. E., and Larson, C. R. 1997. Voice F0 responses to pitch-shifted auditory feedback: A preliminary study. *Journal of Voice*, 11(2): 202-211.

Burns, E. M., and Ward, W. D. 1978. Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, 63(2): 456-468.

Burns, E. M. 1999. Intervals, scales, and tuning. In D. Deutsch (Ed.). *The Psychology of Music* (pp. 215-264). San Diego, CA: Academic Press.

Hartmann, W. M. and Doty, S. L. 1996. On the pitches of components of a complex tone. *Journal of the Acoustical Society of America*, 99(1): 567-578.

Krumhansl, C. L. 1989. Why is musical timbre so hard to understand? In: S. Nielzen and O. Olsson (Eds.). *Structure and Perception of Electroacoustic Sound and Music* (pp. 43-54). Amsterdam: Elsevier (Excerpta Medica 846).

Moore, B. C. J. 2003. *An Introduction to the Psychology of Hearing*. San Diego, CA: Academic Press.

Mürbe, D., Pabst, F., Hofmann, G., and Sundberg. J. 2004. Effects of a professional solo singer education on auditory and kinesthetic feedback – a longitudinal study of singers' pitch control. *Journal of Voice*, 18(2): 236-241.

Ogawa, Y. and Murao, T. 2004. Flat pitch production by accurate singers: Relationship among pitch discrimination, stimulus model, and pitch-interval matching accuracy. In: S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster (Eds.). *Proceedings of the 8$^{th}$ International Conference on Music Perception & Cognition* (Evanston, IL) (pp. 723-724). Adelaide: Causal Productions.

Peters, R. W., Moore, B. C. J., and Glasberg, B. R.. 1983. Pitch of components of complex tones. *Journal of the Acoustical Society of America*, 73(3): 924-929.

Platt, J. R. and Racine, R. J. 1985. Effect of frequency, timbre, experience, and feedback on musical tuning skills. *Perception and Psychophysics*, 38(6): 543-553.

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

Risset, J. C. and Wessel, D. L. 1999. Exploration of timbre by analysis and synthesis. In: D. Deutsch (Ed.). *The Psychology of Music* (pp. 113-169). San Diego, CA: Academic Press.

Russo, F. A. and 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.

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

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

Sundberg, J. 1978. Effects of the vibrato and the 'singing formant' on pitch. *Journal of Research in Singing*, 5(2): 5-17.

_________. 1994. Perceptual aspects of singing. *Journal of Voice*, 8(2): 106-122.

Terhardt, E. 1971. Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. In: *Proceedings of 7th International Congress of Acoustics* (Budapest), 3: 621-624.

_________. 1988. Intonation of tone scales: Psycho-acoustic considerations. *Arch. Acoust.* (Poland), 13: 147-156.

_________. 2000. Pitch shifts and pitch deviations. http://www.mmk.ei.tum.de/persons/ter/top/pshifts.html, 17 February 2006.

Titze, I. R. 1994. *Principles of Voice Production*. Needham Heights, MA: Allyn and Bacon.

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

Worthy, M. 2000. Effects of tone quality conditions on perception and performance of pitch among selected wind instrumentalists. *Journal of Research in Music Education*, 48(3): 222-236.