Eesti Muusika- ja Teatriakadeemia

Allan Vurma

Häälekvaliteet ja helikõrgus laulmisel:
mõningad taju ja moodustusega seotud aspektid

Eesti Muusika- ja Teatriakadeemia Väitekirjad 3
Tallinn 2007

Estonian Academy of Music and Theatre

Allan Vurma

Voice Quality and Pitch in Singing:
Some Aspects of Perception and Production

Estonian Academy of Music and Theatre Dissertations 3
Tallinn 2007

Estonian Academy of Music and Theatre Dissertations 3

Cover design: Liisa-Triin Vurma
Layout: Katrin Leismann
Translation of the introductory section into English: Meelis Leesik

This research has been supported by grants # 3238 and
4712 from the Estonian Science Foundation

Estonian Academy of Music and Theatre
Department of Musicology
Rävala pst. 16
Tallinn 10143

Supervisor: professor Jaan Ross
Opponent: Ph.D. Rytis Ambrazevičius
The date of defence: November, 12, 2007

ISSN 1736-0714
ISBN 978-9985-9797-2-3
Trükitud AS Võru Täht, Oja 1, 65609 Võru

Table of Contents

Publications included in the doctoral thesis

The thesis at hand is based on the following publications, which are referred to in the text by the corresponding Roman numerals:

I: Vurma, A. & Ross, J. (2003). The perception of "forward" and "backward placement" of the singing voice. *Logopedics Phoniatrics Vocology, 28*: 19–28. The journal is available at www.informaworld.com

II: Vurma, A. & Ross, J. (2006). Production and perception of musical intervals. *Music Perception, 23*: 331–344. Copyright © 2006 by the Regents of the University of California/Journal's Sponsoring Society.

III: Вурма А., Росс Я. и Огородникова Е.А. (2006). Восприятие вокальных музыкальных интервалов. *Сенсорные системы, 20*: 117–125.

IV: Vurma, A. & Ross, J. (2007). Timbre-induced pitch deviations of musical sounds. *Journal of Interdisciplinary Music Studies, 1*: 33–50. The journal is available at www.musicstudies.org

Abstract

The thesis at hand comprises four independent research articles, which describe the results of three different studies on the perception and production of singing voice and the terminology of singing training.

The first study, which involved perception tests with synthesized sound stimuli resembling a natural singing voice, revealed that the opposition of the qualities of 'forward/backward' placement, which singing teachers use in characterising the singing voice, can be related to several different acoustic qualities of singing voice. The participants of the test perceived the singing voice as 'placed forward' (as opposed to the quality of 'placed backward'), when (1) the frequencies of the first and the second formant of the sung vowel were higher and/or (2) the frequency of the singer's formant was higher and/or (3) the relative level of sound pressure in the singer's formant was higher.

The second study involved a production experiment in which professional singers were asked to sing three different melodic intervals (a minor second, a tritone and a perfect fifth) both in an ascending and a descending sequence. The results of the study demonstrated that the singers' intonation was characterised by the same regularities that had been described in previous laboratory experiments involving the tuning of sounds generated by a sound generator. The minor intervals (the minor second) were characterized by the tendency of being sung more narrowly and the major intervals (the fifth) more widely than the respective equally tempered values. The tritone, which is relatively rare in musical practice, varied the most, i.e. it was sung the most inconsistently. The participants who had completed a course of musical training were inclined to consider the melodic intervals produced by a singing voice as being correctly intonated even when their deviation from the respective equally tempered values was of a magnitude of 20 to 25 cents.

The third study included a perception test whose participants had to compare the pitches of sounds having different timbres (those of classically trained tenor, piano and oboe) and assess whether the pitch of the second sound was sharp, flat or in tune by reference to the pitch of the first. It turned out that the singing voice was perceived as approximately 20 cents higher than the oboe or piano sound with the same fundamental frequency. Additionally, when the singers were asked to sing the pitches of the oboe and piano sounds that had the same fundamental frequency, the matching sound they produced for the oboe-timbre stimulus was on average six cents higher than that for the piano-timbre stimulus. The reason for this shift in the perceived pitch probably relates to the location of the spectral centre of gravity of the stimulus sound on the frequency axis.

Foreword

The doctoral thesis at hand has been conceived as a part of the author's endeavour to improve his understanding of the various aspects of vocal techniques and of the perception of the singing voice. In addition to purely academic curiosity, the research reported below has been motivated by my long-time professional activities as a singer and a singing teacher. A factor that has been of particular importance in channeling my motivation into actual research is my degree in engineering, which, in addition to my profession as a musician, allows me to approach the problems of singing voice and singing performance in a perspective that is likely to elude performers who have only been trained as musicians.

One of the reasons that spurred me to further my understanding of the different possibilities of using the human voice as a musical instrument and to attempt to shed light on the mechanisms that shape the perception of the listeners lies in the conflicting aesthetic views and requirements that I had to deal with at the beginning of my career. For a beginner vocalist, that conflict made it difficult to see the precise relation between the various vocal techniques and methods and the singing performance that resulted from applying these. At the beginning of the 1980s, the singers in the former Soviet Union, including Estonia, were trained to use their voice mainly to perform the operatic music of the Romantic era, which at the time was suffused with Soviet pomp and pathos. The younger generation, who resisted both the dominant political ideology and the official views on the aesthetics of art, was mostly interested in reviving the early music and folk music traditions.

In those years I was a student of singing at the Tallinn State Conservatory, a bastion of conservative ideology. At the same time, I was working as a singer in the Estonian Philharmonic Chamber Choir, directed by the young and rebellious Tõnu Kaljuste. It did not take me long to realise that without a proper training in singing techniques the exciting and innovative aesthetic aspirations that were being pursued in the chamber choir would remain doomed to amateurism. The training offered at the Conservatory, however, was rather one-sided and rigid and only included a limited number of techniques suitable for implementing aesthetic innovations in a chamber choir. As a result, I realised the need to gain a better overall understanding of the field of vocal methodology, one that would help me to find solutions to situations pitting conflicting principles against one another and to fathom the nature and the underlying causes of their conflicts.

At the beginning of the 1990s, after Estonia regained its independence, I was lucky to meet a number of individuals whose support and help have allowed me to channel my curiosity into research and whom I would like to thank here. One of

them was Dr. Einar Meister, Head of the Laboratory of Phonetics and Speech Technology at the Institute of Cybernetics, who kindly permitted me to use the Institute's laboratory and brought me up to date with the latest research results in phonetics, a branch of science that is highly relevant to singing studies.

I also received significant support from the academician Dr. Jaan Ross, who is also the supervisor of the present thesis, and was then Research Director of the Institute of the Estonian Language. His help in guiding me through the intricacies of the world of science has been invaluable. His continued assistance has permitted me to condense several of my research projects into a Master's thesis, and complete a number of research articles that have been published in international research journals and collections of articles. It is largely thanks to Dr. Ross that I have been spurred to make numerous conference presentations and publish a monograph on vocal methodology.

In addition to Dr. Meister and Dr. Ross, I would also like to thank Dr. Arvo Eek and Mr. Mart Rohtla from the the Laboratory of Phonetics and Speech Technology for their useful insights and many fruitful discussions.

Last but not least, I wish to extend my gratitude to those who have helped me in translating my work into English, including Professor Ilse Lehiste, Professor Lawrence Feth, Mr. Meelis Leesik and Ms. Sirje Ainsaar.

1. Introduction

The doctoral thesis at hand comprises four independent research articles, which deal with various problems related to the perception and production of singing voice in the context of the classical Western music culture. As mentioned above, in one way or another, I have had to tackle these issues as part of my professional activities as a singer and a singing teacher. For this reason, my research has not been oriented only to answering general theoretical questions, but also to solving certain practical problems encountered in the professional activities of musicians. A list of the articles can be found on page 6 above; references to the articles will below be given using the respective Roman numerals.

1.1. On communication between musicians and vocal scientists

Musicians mainly aim to achieve practical results—to produce sounds and to control the properties of those sounds according to the piece of music performed and the musician's artistic vision. Although musicians must necessarily have good technical instrumental or singing skills, the possession of a related theoretical knowledge is commonly neither a conscious goal pursued in the course of musical training, nor a criterion for professional success. Much of the musicians' workload involves musical practice, which aims at perfecting the use of the methods chosen for conveying the musical meaning embodied by a piece to a degree where it would become second nature. Certain genres, e.g. opera singing, also require physical work of an intensity close to that demanded of professional athletes, which means that it would be unrealistic to expect all musicians to be able to engage in the intellectual pursuit of theoretical knowledge in addition to their performance work.

A successful musical performance requires the musician to project, at least during the time of that performance, a sense of conviction and an absolute belief in his/her artistic choices, such that the listeners, too, would be convinced of the suitability of those choices. The task of a musician is thus to create a credible illusion that effectively conveys a musical expression and provides a sense of emotional fulfilment to the listeners. This means that focussing exclusively on an isolated phenomenon or parameter of a musical performance, often unavoidable in the pursuit of scientific research, is something that a musician giving a live performance cannot afford.

Musicians and scientists are often faced with the question of whether the scientific analysis of a problem can, in addition to fundamental knowledge, also provide a practical solution that can be of use to performers, i.e. whether the knowledge gained can help to improve the quality of the musician's performance. Active professional

musicians often adopt a sceptical position in this regard. They find that the problems raised by scientists, as a rule, do not bear a particular relevance from the point of view of the musician and the format of the presentation of research results is often too complicated to understand, because it presupposes a great deal of background knowledge that is not part of the traditional musical training.

Musicians and the scientists studying music and the performance of music are both essentially engaged in one and the same field, yet their motives, goals and methods are different. The goal of scientists is to gather reliable knowledge about the world. In order to be reliable, knowledge must command a high degree of probability of being true, which will be the case if it has been acquired by employing the scientific method. The essence of that method lies in collecting empirical evidence, using logical reasoning to analyze it and espousing a generally sceptical attitude. This means that, if the data supporting a new explanation appear conclusive, the scientist will always be prepared to revise his/her own earlier positions and to call into doubt the opinions of previous authority figures.

The requirement of ensuring reliability, however, complicates scientific investigations considerably. Finding a valid methodology (i.e. one that is suitable for finding the answers to a particular question) may not be an easy task. For example, in order to describe a dependency relation, or any other kind of relation between phenomena, one must see to the 'internal validity' of the experiment that is to be carried out. This means that changes in the observed property of the subject matter of research should only be triggered by the manipulation of the independent variable and not by an external factor that is beyond the control of the investigators. The nature of the objects or phenomena investigated, however, is often so complex as to render the isolation of the influence of a particular factor impossible.

On the other hand, when we succeed in creating, in a laboratory setting, nearly ideal conditions in which we are able to modify one parameter or the object or phenomenon and observe a resulting change of some other parameter, the result will often be alien to real life conditions, thus damaging the 'ecological validity' of the experiment. In such a case we will have obtained results which are valid in the 'sterile' environment of the laboratory, but are of not much use for understanding everyday situations in the real world.

For the reasons described above, it may be that instead of studying those problems that are truly relevant and important to musicians, we wind up investigating that which is methodologically easier. The situation resembles that evoked in a well-known joke about a drunk who lost his keys in a dark corner of the street, but started looking for them away from the corner under a streetlamp, since, although it was impossible for the lost keys to be there, looking for them was certainly easier in a spot where there was more light.

The reason why research results sometimes remain useless for raising the actual quality of performance of a piece of music may lie in the use of a research method that employs a restricted experimental setting specifically created in order to ensure

the reliability of the experiment and differs to a smaller or greater extent from the real performance situation. The use of ecologically valid research methods, which imitate real life situations, however, often results in a conflict with the strict criteria of internal validity of an experiment. If preferred, such methods may weaken the reliability of the research results obtained. Thus, while planning research we must each time evaluate which set of criteria (either those oriented to ensuring internal validity or those aiming for ecological validity) are more important to us in order to achieve the desired result, and select the suitable method accordingly.

Moreover, if what we desire is to acquire highly reliable knowledge, we cannot depend solely on the intuition-guided theories and views of the musicians themselves, regardless of the strength of their conviction in regard to these. After all, having run the gauntlet of the scientific method, many common-sense theories and views have unfortunately proved to be inadequate, imprecise or only valid in a limited or a figurative sense. Hence, although notions that are only based on practical experience may be adequate to solve a particular problem, they are likely to fail when used for providing an account of a broader phenomenon.

As an illustration to the above, allow me to briefly evoke an experiment conducted in the piano factory of Leningrad (now, St. Petersburg) at the end of the 1970s (Galembo & Askenfelt, 2003). The teaching staff at the Piano Department of the Leningrad Conservatory were asked to evaluate the sound quality of three different concert pianos (Steinway, Bechstein and a piano built in the local factory) by playing a short piece of their own choice on those instruments. The experts were also asked whether they were able to recognise the piano brand solely on the basis of its sound. All 12 experts responded in the affirmative. A 'blind' test was then conducted: a curtain was used to partition the room so that the experts did not see the instruments, after which scales, chords and arpeggios were played on them. The experts were in each case asked which particular instrument of the three was being played. Contrary to the experts' subjective opinion of their competence, none of them was able to provide reliable judgments regarding the identity of the brand played.

In the next stage of the experiment, three pianos were positioned in a circle, their keyboards facing inwards, so that it was possible to play all of them while sitting on a revolving stool in the middle. The same group of experts were blindfolded and seated on the stool, which was then spun so that the experts would not be able to tell which of the pianos they were facing. Once more, the experts were asked to identify the brand of the piano—without being able to see it, but now allowed to play themselves while listening to its sound. Surprisingly, this time the number of mistakes was insignificant: over eighty percent of the answers were correct. Thus, it was indeed possible to identify the pianos with sufficient certainty, mainly by the difference in the sensation obtained by hitting the piano key and feeling its mechanical response, as opposed to the difference in the sound itself. The kinesthetic sensation from the keys had become interfused with the perception of the sound timbre and

the pianists/experts were unable to distinguish one from the other. The use of the scientific method showed that the pianists' widespread and intuitively obtained belief regarding the quality of different piano brands was in part erroneous. The fact that the pianos produced by different companies were indeed different turned out to be true. However, the difference was not in the quality of the sound but in the mechanical reaction of the keys of the pianos to the touch of the pianist's fingers. A fallacy of this type would probably cause no problems in a context in which the pianists communicate only between themselves. The experiment was, however, initiated by piano builders who were trying to improve the actual sound qualities (e.g. by improving the manufacturing technology of soundboards and strings) of the pianos produced in the Leningrad factory, in order to follow up on the frequent complaints by the pianists. As it turned out the crux of the problem was in the mechanic components of the keyboard. This realisation was, however, only made possible as a result of conducting the experiment described.

1.2. On the history of applying scientific knowledge in vocal methodology

Although musicians, including singers, are often sceptical of scientific research, the history of vocal art and its teaching still demonstrates that several methods and vocal techniques, present or past, are based on information and notions that are rooted in knowledge obtained via scientific methods. At the same time, the body of scientific knowledge is constantly growing and many past theories or notions have later turned out to be imprecise or false. Some of the explanations based on notions that have later been refuted have once also been used in vocal pedagogy and are sometimes still in use. Yet, not all singers and singing teachers interested in scientific knowledge and its application have not always been able to completely understand the meaning of the respective knowledge. Because of this, we can find pseudoscientific theories and teaching methods both from the past and the present of vocal methodology. The concepts, explanations and notions of the workings of the vocal apparatus as used and understood in the practice of vocalists may not always be compatible with the treatment of the respective topics by scientists.

The history of the scientific study of matters relating to vocal pedagogy and the human voice extends back to ancient times. However, a closer interaction between the two fields can only be observed from the middle of the 19th century onwards. Treatises on vocal pedagogy produced before the 19th century were mainly confined to the problems of aesthetics and style. For example, Isidor de Seville who lived in the 7th century writes in his "Etymologies" that the perfect voice should be *suavis, alta* and *clara*[1] (Dyer, 2000). In the 1280s, Jerome di Moravia, a Dominican priest,

[1] Latin for 'sweet, high and clear'—transl.

described three intuitive categories of voice: *vox pectoris, vox guturris* and *vox capitis*[2], which probably correspond to different registers of voice or parts of the vocal range in the modern sense (ibid.).

Vocal science is considered to have been founded by Claudius Galen who wrote an essay on the human voice in the 2nd century. He gave a description of the human throat and realised the role of the brain in the production of voice (von Leden, 1997). Starting from the end of the 17th century, significant progress in studying the mechanisms of voice production was made by French scientists. For example, Claude Perrault (1613–1688) compared the voice organ and its working principle with the flute (ibid.) in his "Œuvres diverses de physique et de mécanique"[3]. Denis Dodart (1634–1707) introduced the concept of 'voice lips' (translated as 'Stimmlippen' into German, it is still used sometimes in Germany) in his article "Mémoire sur les causes de la voix de l'homme"[4] at the beginning of the 18th century. He found that the pitch of the voice depends on the tension of the 'voice lips' and thought that the trumpet is a better analogy for the voice apparatus than the flute, since the 'voice lips' function similarly to the lips of a trumpet player in producing a tone (ibid.). Antoine Ferrain published a series of articles on the results of experiments conducted on the throats of animal and human cadavers in 1741. He used the term 'cordes vocales' (vocal cords) and correctly observed that the pitch of the voice depends on the frequency of vibration of the vocal cords (ibid.). Albrecht von Haller, a Swiss who worked in Germany, wrote about the resonation of voice in the nasal and oral cavities and the paranasal sinuses in his "Elementa Physiologiae"[5], published in 1761 (ibid.).

By the time the scientists of those days had reached these conclusions, largely accepted even today, the art of opera singing had already existed for more than one hundred and fifty years.[6] The secrets of the art of singing were passed on as they are today—in the direct interaction between teachers and students throughout long years of study. Experience and techniques were acquired by trial and error, a process in which intuitive instincts played a major role. Not much was written on the subject of teaching opera singers. There are only three better-known authors from the 17th and the 18th centuries. The first of these, Giulio Caccini, a singer and a singing teacher, was a member of the *Firenze Camerata*, an eminent group of Florentine humanists, musicians, poets and intellectuals, which, among other things, is known in history for laying the aesthetical foundations of the new art form of opera. Caccini described the aesthetics of the new vocal style in the foreword to the song collection "Le Nuove

[2] Latin for 'chest voice, throat voice and head voice'—transl.
[3] French title, freely translated as 'Various Works on Physics and Mechanics'—transl.
[4] French title, freely translated as 'Treatise on the Causes of the Human Voice'—transl.
[5] Latin title, freely translated as 'Principles of Physiology'—transl.
[6] In the history of music, the first work that can be called an opera is Jacopo Peri's 'Dafne' (1597).

Musiche"[7] published in 1601 (Mori, 1953). Besides Caccini, there were also two famous castrato singers/singing teachers who wrote more substantial treatises on the subject. Pietro Tosi, who lived and worked in Bologna, expounded his ideas in "Opinioni di cantori antiche e moderni o sieno osservazioni sopra il canto figurato"[8], published in 1723, and Giambattista Mancini, the singing teacher at the Imperial Court of Vienna, wrote a work called "Pensieri e rifflessioni pratiche sopra il canto figurato"[9] in 1774 (ibid.). As the respective titles indicate, these treatises mainly dealt with stylistic techniques for singing embellishments, although Mancini also describes the effects of mouth position on the voice (ibid.). However, these first attempts were a far cry from providing a systematic description of the theoretical foundations of vocal techniques.

Singing teachers started to show an interest in science in the middle of the 18th century and mostly in France. One of the first works reflecting this was "L'art du chant"[10] (1754) published by Jean-Baptiste Bérard, one of the best-known singers and singing teachers of the time. Bérard explained the basics of breathing for singing in accordance with the scientific standards of the time and described the position of the throat on the basis of actual observation data (Newton, 1984: 61). A revolution in singing training was brought about by the activities and works of Manuel García II (1805–1906), one of the most legendary singing teachers who worked in Paris and later in London. He grew up in a family of famous singers. For instance, Giaccomo Rossini wrote the part of Count Almaviva in the "Barber of Seville" for his father, a famous tenor who also bore the name of Manuel García. Later, the mezzo-sopranos Maria Malibran and Pauline Viardot, sisters of Manuel García junior, also became famous (Sell, 2005: 9–39).

It often happens that significant paradigm changes in some areas are triggered by random incidents in the lives of the people concerned. The life of the young Manuel García was also largely shaped by an accident. Having studied singing under the guidance of his father, Manuel embarked on an extensive opera tour in America together with his family when he was twenty years old. Probably due to a lack of experience and an excessive workload he lost his voice and gave up singing when he was only 24 years old. Having returned to Europe, for a short period Manuel worked in a military hospital, where he had the opportunity to study the anatomy of the throat by examining soldiers with neck wounds, and contemplated the reasons for his own failure as a singer. His research was evidently most fruitful, since having soon returned to singing, though now as a singing teacher, his success was such that he was soon appointed Professor of Vocal Music at the Conservatory of Paris. Manuel was only

[7] Italian for 'New Music'—transl.

[8] Italian title translated as 'Opinions of singers ancient and modern, or observations on figured singing'—transl.

[9] Italian title translated as 'Practical reflections on the figurative art of singing'—transl.

[10] French title translated as 'The Art of Singing'—transl.

thirty years old at the time (ibid.). In spite of his success as a teacher, Manuel García junior is primarily remembered in history for his "Traité complet de l'art du chant"[11] (García, 1974), originally published in 1841, and his lecture at the Paris Academy of Sciences, held one year before the publication of the book. García gave a detailed description of the anatomy and mechanisms of the vocal apparatus and regarded this knowledge as indispensable for both singers and singing teachers. In his presentation at the Academy of Sciences García, assisted by his students, described and demonstrated two different voice timbres: *timbre clair* or 'clear' timbre, which was faithful to the aesthetics of the so-called Old Italian school, and *timbre sombre* or 'covered' timbre, which permitted to achieve a totally new manner of expression and a strong carrying power of the voice. According to some sources, he also invented the laryngoscope in 1855 (fourteen years later) and was the first person to use this instrument to observe his own vocal cords producing a sound (ibid.).

Although the influence of Manuel García II on the development of vocal pedagogy is undeniable and the use of objective scientific knowledge in singing teaching has gradually increased since his contributions, not all of his contemporaries approved of García's ideas, disagreeing with his position that scientific knowledge was necessary for singers. One of García's opponents during his lifetime was Morell Mackenzie, an otolaryngologist much respected among singers. Although he considered the knowledge of anatomy useful for singing teachers, he regarded the attempt to teach singing by studying the anatomy of vocal organs with a laryngoscope plainly absurd, and compared it to teaching a painter to paint by inviting him to study the anatomy of the eye with an ophthalmoscope (ibid.).

After the publication of the treatises of Manuel García II, singing teachers started to show a gradually growing interest in the scientific treatments of vocal production. By the middle of the 19th century almost half of the published textbooks on singing contained chapters on the anatomy of the voice apparatus, and by the end of the century all relevant publications contained such information (ibid.). Several treatises with relevance to singing training were written by laryngologists and scientists. For example, the laryngologist Louis Mandl published his article "De la fatigue de la voix"[12] in the French medical journal *Gazette Médicale*[13] in 1855, describing in it the advantages held for singers by low/abdominal breathing (ibid.). Before that, singers had been using a 'noble' posture (with the chest lifted). As the result of the article, several artificial means and techniques came into fashion in order to acquire 'low' breathing.

The progress made in acoustics also played a significant role in the development of vocal pedagogy. There were two competing theories of voice production in the second half of the 19th century (Fletcher, 1929). According to the 'harmonic' theory, the

[11] French title translated as 'A Complete Treatise on the Art of Singing'—transl.
[12] French title translated as 'Of the Fatigue of the Voice'—transl.
[13] French title translated as 'The Medical Gazette'—transl.

vocal cords create a complex sound, which consists of a fundamental and a number of partials. When this sound passes through the vocal tract, which acts as a resonator, the partials that are closer to the resonant frequencies of the vocal tract are amplified to a greater magnitude than others.

The proponents of the second, 'inharmonic' theory, thought that the glottis only produces puffs of air, and that these puffs, when they enter the vocal tract, excite the sine waves corresponding to the resonant frequencies of the tract. The puffs of air need not follow each other at a frequency that equals the resonant frequency of the tract and they do not consist of partials having different frequencies. The advocates of the harmonic theory, which in its main principles is still valid today, included Charles Wheatstone as well as one the most famous German scientist of the time, Hermann Helmholtz. The second position, by now refuted, was defended by R. Willis, L. Hermann and E.W. Scripture.

Both theories were soon also discussed and debated by singers and singing teachers. For example, Enrico Delle Sedie, an Italian baritone known for his parts in Verdi's early operas, published a book called "L'estetica del canto e dell' arte melodrammatica"[14] (printed in Paris), in which he presented a vowel diagram drawn up according to the theory of Helmholtz. Unfortunately, singers were also inspired by the inharmonic theory, which has been refuted by now. For instance, E.G. White in his "Science and Singing" (1909) elaborated a theory of the resonance of paranasal sinuses (once popular also in Estonia), according to which the resonance related to the carrying power of the voice emanates from the maxillar and the frontal sinuses (Sell, 2005: 35). Another well-known book is "Meine Gesangskunst"[15] by Lilli Lehmann, one of the best-known German sopranos of the 19th century. This book, published in 1902, is mostly known for its fan-shaped diagrams (Lehmann, 1981). Lehmann was convinced that in order to produce a better resonance the singer should 'direct' her or his voice into different cranial and facial regions depending on the pitch of corresponding note (ibid.). Jean de Reszke (1850–1925), a renowned Parisian tenor of Polish origin who exerted a significant influence on the French vocal pedagogy of the time, also recommended 'singing into the mask' (Sell, 2005: 35). That, and his other writings and theories, has posthumously earned him considerable criticism for being one of the reasons why France has not produced (with only a few exceptions) any great opera singers in the 20th century (ibid.).

During the 20th century the body of scientific knowledge on the workings of the voice has increased significantly. Several legendary teachers, who were at the same time brilliant singers, were directly involved in research and raised public awareness of the importance of using scientific knowledge in vocal pedagogy. To note a few most important ones, I should probably mention the Americans William Vennard

[14] Italian title freely translated as 'The aesthetics of singing and melodrama'—transl.
[15] German title translated as 'How to sing'—transl.

(1909–1971), Ralph Appelmann (1908–1993), Richard Miller (1926–) and Jo Estill (1921–). In 1970s, Johan Sundberg, a Swede, studied and described the formation principles of the singer's formant, which increases the carrying power of the voice. About ten years earlier, Gunnar Fant, another Swede, had earned international recognition for his theory of the acoustics of speech production (Fant, 1960). Sten Ternström, who replaced the legendary Johan Sundberg as the head of the music acoustics group (one of the world's best) at the Royal Institute of Technology in Stockholm, has studied and described various acoustics problems encountered in choir singing. Ingo Titze, the leader of the research group at the University of Iowa in the United States, has among other things channelled his efforts into describing the different working modes of vocal cords. He often writes popular science articles for the Journal of Singing in order to bring his research results closer to an audience of singers, mainly focussing on the aspects that may be of interest to it.

2. An analytical summary of the aims and of the substance of the research conducted by the candidate

In terms of singing technique, the singer must, while singing, constantly engage in shaping and adjusting three acoustic parameters of his/her singing voice. These are timbre, pitch, and sound level. The articles that constitute the doctoral thesis at hand deal with matters pertaining to the first two. The studies that form the basis of these articles have focussed on building bridges between, on the one hand, the theories and approaches employed by vocalists and singing teachers and, on the other hand, the systematic knowledge acquired by applying modern scientific methods to describe the functioning of the vocal apparatus and the mechanisms of human perception. Building such a bridge appears particularly desirable in connection with the terminology used in teaching singers, since the terms used by singers in conveying their practical knowledge may often differ from those employed by phoneticians or phoniatrists in describing the same phenomena. It is also important in connection with topics related to pitch, both for vocalists and, in a broader fashion, for the performance of music generally. After all, it is often the case that the musician's understanding of the topics related to intonation resembles an idealised abstract scheme and fails to take into account certain facts the nature of which can only be elucidated through systematic research.

Singing and the singing voice can be investigated either by focussing on the production in human vocal apparatus of sounds with various acoustic properties, or on the aspects of the perception of such sounds and sound groups. The articles that have been included in this thesis employ an integrated approach that attempts to link both production and perception into a single comprehensive description. In fact, such an intertwining of the two spheres is also characteristic of the real singing situation, from the vantage point of the singer—in a live musical performance, the shaping of the acoustical properties of voice such as pitch, sound level and timbre proceeds on the basis of an ideal sonic image created in the singer's mind the musical work to be performed. One of the components of this process is feedback through auditory perception. Hence, the production and perception of a sound form an integral whole also for the singer him/herself.

The thesis at hand consists of three separate studies, which will be dealt with in detail in the following section. First, subsection 2.1. will focus on topics related to the perception and production of, as well as the terminology used to describe certain aspects of timbre. These are treated in article I. Then, subsection 2.2. will deal with the two remaining topics, both of which relate to the various problems connected to

producing and perceiving the pitch of sounds. The relevant investigations and their results are set out in articles II, III and IV.

2.1. Perceptions of the 'forward' and 'backward' placement of singing voice

Many methods and notions of singing training, together with the explanations that go with these, are still passed down from teacher to student in direct and individual coaching, sometimes without recalling or knowing their origins. In addition to those that correspond to modern scientific knowledge, this way of pedagogical communication may convey theories and techniques based on dated scientific approaches that have either been proved wrong or whose substance has been modified considerably. It is also possible that the teacher imparts to the student understandings based on metaphors and subjective (bodily) perceptions. For this reason, it is likely that the meaning and content of certain terms and methods used in singing training will be difficult to fit into the contemporary scientific paradigm. It is important to realise this, provided one wishes to communicate beyond the bounds of his/her immediate speciality.

The aim of the first article of the doctoral thesis at hand is to provide a description of the acoustical nature of the qualities of the singing voice that correspond to one such term pair, that of 'forward/backward' placement and the possible ways of producing those qualities. The terms started to be employed probably in the second half of the 19th century and may have been inspired by the previously described inharmonic resonance theory (Vennard, 1967: 120).

The introductory part of the article describes the reasons that have led to the use in singing training of explanations and concepts the content of which does not correspond to the insights achieved by contemporary science. In what follows we provide a summary of previous research that describes the relationships between the acoustical parameters of singing voice and the various adjectives used to describe the timbre of that voice. The method that we employed in our study was conducting expert assessments of the correspondence of perceptions of vocal stimuli to the qualities of 'forward' or 'backward' placement of the singing voice. The respective stimuli, resembling a natural singing voice, were generated on a computer and their parameters were modified systematically in accordance with the plan of the experiment. Expert participants included both professional singers and singing teachers from Estonia and abroad. Foreign experts provided their assessments through a website that had been set up specifically for that purpose. The study described in the thesis at hand was a follow-up on earlier research (Vurma & Ross, 2002), in which the investigators focussed on the production of singing voice corresponding to the qualities of 'forward' and 'backward' placement by live singers.

The results of the study show that the perception of the singing voice as placed 'forward' or 'backward' may depend on several relatively independent acoustical

parameters. The parameters in question are (1) the frequency of the first and the second formant, related in perception to the quality of the vowel, (2) the frequency of the singer's formant, important for classifying singers by voice categories and (3) the relative level of the singer's formant in the spectrum of his/her voice.

In the course of the research described in the article, the investigators were faced with the need to resolve a conflict between the ecological and internal validity of the method chosen, much alike to what was described in section 1.1. of the introductory chapter above. The use of sounds recorded by live singers strengthens the method's claim to ecological validity, but at the same time weakens its internal validity, because the acoustical parameters of the singing voice produced by a live singer are never completely stable. The use of stimuli recorded by singers also represents a problem for the reason that the singer is normally unable to modify a single parameter of his/her voice exclusively, maintaining other parameters at fixed values, an aspect important from the point of view of ensuring a high internal validity of the experiment. In the case of the study at hand, the investigators opted for an increased internal validity and used computer-generated vocal stimuli resembling the singing voice. The results of the study complemented those of the previous research mentioned above (Vurma & Ross, 2002), in which the accent was placed on ecological validity, and the use of sounds produced by live singers was given preference. In fact, certain compromises were also made in the study described here. Thus, in some cases the authors derogated from strictly adhering to the criteria of internal validity—in order to ensure a better likeness to the human voice, the parameter values of the vibrato of the stimuli were adjusted in addition to modifying the frequencies of their formants.

2.1.1. The significance of the results obtained

The importance of the research outlined above consists in the fact that it provides a multidimensional account of the concepts of 'forward' and 'backward' placement. This permits to understand the grounds of communication problems that are apt to arise in situations in which different users of these concepts focus on diverging and narrowly delimited aspects of the underlying quality of singing voice. The study also described the possible articulatory means the employment of which was likely to lead to the production of the quality of 'forward' placement in the place of that of 'backward' placement, which is generally used with a pejorative connotation.

The results of the study suggest that, similarly to the term pair 'forward' and 'backward' placement, there may be other analogous terms that are used in teaching singers and that are characterised by a semantic field that accommodates multiple interpretations. The specific referents of such terms are determined by the context of their use and may depend on the experience and competence of the user. Since, in the training of singers, the practical acquisition of vocal techniques and skills required in the work of a vocalist is of primary importance, while the ability to grasp the acoustic, physiological, anatomical or articulatory mechanisms of these, albeit recommended, will not be regarded as strictly necessary in order to acquire a good

singer's skills, the lack of precision, in scientific terms, of the terminology or explanations used during training may not result in any communication problems during singing training or impede the communication between singers. The singers or singing teachers will, however, find it difficult to make themselves understood when they try to communicate beyond the modalities of their immediate speciality. A case in point is, for instance, a situation in which a singer would need to exchange information regarding the singing voice and its production with a phonetician. Unlike that of the singers', the latter's use of the terminology describing phenomena related to singing will, as a rule, be grounded in the contemporary scientific paradigm.

2.2. Articles dealing with topics related to pitch

Articles II, III and IV of the doctoral thesis at hand deal with the problems related to producing and perceiving sounds, and combinations of sounds (musical intervals), of varying pitch in contexts that involve singing and the singing voice. In music, the pitch of a sound has been regarded as having primacy over the properties of duration and timbre. The reason for this is that the removal from a musical work of information regarding the pitch of its component sounds generally renders the work unrecognisable. At the same time, the modification of, for instance, the duration of those component sounds or, in a less severe attempt on the integrity of the work, the modification of their timbre does not render the identity of the work nearly as opaque. In singing, the justness of pitch is somewhat more difficult to ensure than in playing most of musical instruments, because a singer cannot resort to auxiliary sensations such as those that assist string players (the position of the hand pressing the strings against the fingerboard) or brass players (a combination of the instrument's valves). For example, the overall falling or rising of pitch during the performance of a piece by a semitone or even more is something that occurs quite often even with professional singers performing *a cappella,* but will be highly unlikely in the case of a string quartet. Intonation problems are frequently encountered in ensemble or choir pieces that contain complicated modulations and are performed without instrumental accompaniment. In order to be able to foresee the possible intonation problems and to avoid slip-ups, it is important to understand the factors that influence intonation. It is also important to possess information about the actual intonation encountered in live performances of musical works and about the way that the listeners perceive performances that exhibit one or the other intonation pattern. For instance, interviews with musicians often show that they have an idealised and by far unrealistic perception of the 'purity' of intonation attained in live performances, as well as of their own ability to assess it.

The aim of the articles treating pitch-related issues in the thesis at hand relates to two rather clearly delimited fields. Articles II and III report the results of studies conducted to investigate the ability of singers to intonate various musical intervals without accompaniment. The studies also focussed on the ability of listeners to identify

the possible deviations of these intervals from theoretical values and the eventual presence of certain systematic deviation tendencies in intonation. The aim of the study reported in article IV is to describe the effect of timbre-difference on pitch in a comparison of two sounds.

Most musical works in the Western culture as well as in other cultures are based on musical scales built up of a specific number of sounds, each of which has a discrete pitch. These scales are organised on the basis of an underlying principle (e.g. equal temperament, Pythagorean tuning or just intonation). According to the definition of the American Standards Association, pitch is an attribute of sounds that permits to organize sounds into a musical scale (ASA, 1960). Hence, the determination of the pitch of a sound requires a subject, i.e. a person who makes the ranking assessment. Subjective assessments, however, are to a larger or smaller extent unstable and uncertain, for they depend on the particular characteristics of the person giving the assessment and the context in which the assessment is made. Although the pitch of a musical sound is usually closely related to its objectively measurable fundamental frequency, or the iterative frequency of its wave period, that pitch is also influenced by a number of other factors such as the level and the spectrum of the sound (Terhardt, 1988; Rossing, 1990: 109). Thus, for instance, when the level of a low and quiet sine sound with a fundamental frequency that is below the boundary of about 2000 Hz is increased, it will be perceived as flattening in pitch, whereas a similar adjustment to sounds above 2000 Hz in frequency will result in their pitch being perceived as higher. The magnitude of this effect depends on the perceiver and may reach values of five to ten percent[16], although magnitudes of one to two percent are the rule (Plack & Oxenham, 2005: 9).

A special case, in which the simple logarithmic relationship between the fundamental frequency of a sound and its pitch does not apply is that of auditory illusions. For instance, in the case of the tritone paradox it will be impossible to determine definitively whether, given a particular configuration of the sound spectra, an interval is rising or falling. Likewise, in the case of the socalled 'Shepard illusion', the listener perceives an endlessly rising scale, although the parameters of the sounds succeeding one another are in fact repeated cyclically (Deutsch, 1999) (for the results of previous investigations of the relationship between timbre and pitch, see also article IV, pp. 34–36). In music practice, the pitch of a sound is often determined by, and instruments tuned with, the help of electronic devices that function by measuring the fundamental frequency of sounds; tuning forks that produce a sine wave of a set frequency (usually, 440 Hz, corresponding to the A above middle C) are also common. When such tuning aids are used, other external factors that are susceptible to slightly shift perceived pitch are automatically disregarded. This is apt to raise

[16] The change of approximately six percent in the frequency of the fundamental corresponds to a semitone interval.

questions such as whether only considering the frequency parameters of sounds may lead to problems in tuning and in intonation. To answer these, it is important that we are able to assess the magnitude of the impact of such external factors and the situations in which they are manifested.

The other set of problems has to do with the principles for selecting a discrete pitch level out of an ungraded continuum of a musical scale. It also regards the precision that musicians are able to attain in matching the selected pitch, as well as the deviation margin within which the listeners will consider such a pitch or a musical interval 'in tune'.

Normally, descriptions of musical scales are given by reference to the fundamental frequency values of their elements and to the ratios of those values. This means that the possible effects of sound level and of timbre on perceived pitch are disregarded. In the case that other properties that characterise a sound apart from its fundamental frequency value (such as sound level and timbre) are relatively stable, such an approach should not create problems (in effect, sound level and timbre have not been taken into account in the studies reported in articles II and III). The scales used in different cultures may differ to a greater or lesser extent. Carterette (1999: 734) has proposed four conditions that should be met by an enduringly useful musical scale. These can be summarised as follows: (1) the pitches of the adjacent tones of the scale should be separated from one another by a sufficient distance in order to allow listeners to distinguish between them easily; (2) the sound with a frequency f should resemble the sound with the frequency $2f$ and the sound with that of $f/2$, since the distance between the respective sounds equals an octave; (3) the scale should consist of approximately seven constituent elements; (4) all intervals of the scale should be integer multiples of the minimum interval. These conditions are met by the equally tempered diatonic scale used in the contemporary Western music culture. The third condition, the requirement of approximately seven elements, emanates from the results of an experiment performed by Miller (1956), who showed that subjects are only able to maintain consistency in classifying stimuli in a one-dimensional psychophysical continuum into five to nine (averaging seven) categories.

The view of the equally tempered musical scale as the best possible solution, however, is difficult to reconcile with another tenet of Western musical theory, according to which our perception tends to give preference to certain 'natural' intervals, the ratio of whose frequencies can be described by reference to that of small integers. Thus, for instance, the natural octave corresponds to the frequency ratio of 1:2, the fifth to that of 2:3 and the fourth to 3:4. The smaller the ratio, the more consonant, or musically harmonious, the corresponding pair of sounds will be. In equally tempered tuning the only natural or acoustically just interval is the octave, whereas the values of all other intervals deviate somewhat from the ratio of the relevant pair of small integers.

Various explanations have been advanced in support of the preference of natural intervals. If the fundamental frequencies of two sounds relate to one another as two small integers, the frequencies of a significant number of the partials of these sounds

will coincide—in the case of an octave, every second, in the case of a fifth, every third and in the case of a fourth, every fourth. If there is a small discrepancy between the frequency ratio and the ratio of the relevant pair of small integers, the potentially coincident partials will exhibit a small discrepancy, too. This, in turn, will result in an amplitude modulation of the composite sound, with a speed equalling the difference between the frequencies of the corresponding partials. Such modulation is perceived as the roughness of the interval, or beating[17].

According preference to harmonies based on the ratio of small integers may be related to the general mechanisms underlying the perception of complex sounds. Contemporary researchers have advanced a number of theories all of which include a pattern matching model (Goldstein, 1973; Terhardt, 1974). Humans have the ability to recognise a variety of patterns and, if a part of a known pattern is missing, we tend to supply it from our memory. Our hearing appears to expect complex sounds to consist of harmonic partials also when this is not exactly the case.

The human auditory system first analyzes each partial of a complex sound separately and then attempts to fit the results into a model consisting of harmonic partials. The sound is thus attributed a pitch that provides the best possible fit with the model. According to Goldstein's theory, the vibrations of the basilar membrane of the snail in human internal ear do not peak sharply but fluctuate around a statistical mean. In Terhardt's view, in perception the position of a partial is, to a certain extent, influenced by the partials surrounding it, which causes the processing 'template' in the brain to stretch and may result in a shift in the pitch of the sound.

The third explanation is based on the assumption that the brain 'prefers' a combination of frequencies in the case of which the pattern of the nerve impulses represents a common periodicity (Patterson, 1986). Indeed, such a situation occurs precisely when the frequencies of sounds produced simultaneously correspond to the ratio of the relevant pair of small integers.

Explanations informed by a preference of the ratio of small integers, based on the views stressing either sensory consonance or the common periodicity of the pattern of nerve impulses presume the simultaneousness of the sounds concerned. There have been attempts to expand both of these views to include melodic intervals. Thus, Wood (1961: 181) has advanced a theory that observes that early music was performed chiefly in sonorous caves that were characterised by a long reverberation time. Played in such conditions, the constituent sounds of a melody would have been perceived as virtually simultaneous. In fact, the occurrence of a 'neural reverberation' has been suggested as part of a nerve impulse theory (Boomsliter & Creel, 1971; Roederer, 1973: 145–149). This could provide the foundation for using the ratio of small integers as an intonation criterion also in the case of sounds produced consecutively.

[17] Beating will not be perceived in a sound that is produced with a frequency vibrato, such as that in an opera singer's voice, since the presence of the vibrato obscures beating.

In effect, we can find musical scales that also prefer consonant intervals (the octave, the fifth and the fourth) and that are analogous to the Western diatonic and chromatic scale, in the three large non-Western music cultures—the Indian, the Chinese and the Persian-Arabic. Still, in other musical cultures of the world one can encounter scales in which the frequency ratios of intervals differ considerably from the ratios of small integers. For instance, the adjacent steps of the equiheptatonic scale are separated from one another by a distance of 171 cents, and those of the equipentatonic scale by 240 cents. The equiheptatonic scale can be encountered frequently in South-East Asia, while musical scales consisting of five steps separated from one another by equal distances are in use in the islands of Bali and Java, where some of the most popular musical instruments are the xylophone and the bell chime.

The sound spectra of these instruments, unlike the spectra of most instruments used in Western music, are characterised by inharmonicity, i.e. the lack of partials precisely matching the integer multiples of the sound's fundamental frequency.

The presence of consonant intervals such as the octave, the perfect fifth and the perfect fourth, which are based on the ratio of the corresponding pairs of small integers, in almost all musical cultures of the world seems to indicate that in the course of the historical development of musical intervals, preference was given to these for some natural reason, be it the perception of beating or the common periodicity of the pattern of nerve impulses. Still, studies have shown that beating is largely devoid of significance in the intonation of musical pitch by live musicians. Beating may become significant only in the case of certain musical styles, such as the so-called *barbershop*-style ensemble singing, in which vibrato is avoided and tuning precision may reach two or three cents (Hagerman & Sundberg, 1980), a value that is superior to the pitch difference limen[18] in melodic intervals. In actual musical performance, significant deviations from the standard fundamental frequencies of the elements of the scale used tend to be the norm, regardless of the underlying theoretical tuning system.

For example, the study conducted by Frances (1987) found that the values of the intervals in a 46-sound melody performed by three singers varied within a range of 104 cents (the interquartile[19] range of the same intervals was 38 cents). Still, musically experienced listeners judged that the performance of all three singers had been sufficiently in tune. Ternström and Sundberg (1988) studied the intonation produced by six choir singers performing a cadence of eight tones and found that the standard deviation of the intervals was 13 cents, at the same time recording individual pitch shifts

[18] Typical values of the pitch difference limen in laboratory conditions have been suggested to have a magnitude of 5 cents (Terhardt, 1988). As a rule, it is measured at a certain degree of probability (for instance, 0.7), which means that a participant will be able to correctly recognise the difference in seventy cases out of a hundred.

[19] Inter-quartiles are used to highlight the range of values that is the most representative of the distribution of data. They are obtained by dividing the values of a body of data into four quarters. An interquartile will include the data of the two middle quarters.

of up to ±45 cents. Ward (1970) summarised the results of various studies of violin music by pointing out that in one and the same performance, intervals could vary by as much as 78 cents (for an interquartile shift range of 38 cents). Moreover, the actual intonation did not approximate either just or Pythagorean tuning and did not manifest significant differences in solo as opposed to in ensemble play.

In certain cases, still, pitches that nearly match those of the theoretical values of the tuning system may be produced. Thus, for instance, Kopiez (2003) reports a case of intonation by trumpet players, who in a four-part slow-paced exercise performed the score of the highest part without allowing the pitch to shift more than an average 4.9 cents (standard deviation was 6.5 cents) from the equally tempered values in relation to the bass part. All deviations were within a range from –12 to +23 cents. To practice their parts of the score to a precision-tuned accompaniment, the trumpet players were given 10 days. However, even after several days of practice, they were unable to adjust the tuning of their instruments to that of the accompaniment when three lower parts of the score were provided in just tuning instead of the equal temperament. Kopiez described this as a 'burn in' habit—the result of years of practice based on equally tempered tuning as opposed to just intonation.

Sundberg *et al* (1995) have studied the intonation of singers and the perception of that intonation by ten qualified experts on the example of the performances of Schubert's Ave Maria available on 10 commercially produced CD albums. In spite of the fact that the singers were internationally acclaimed professionals and that the performances had been recorded by experienced sound engineers, shifts of ±40 cents from the values of the steps of an equally tempered scale derived from the harp or piano accompaniment of the performances were common. At the same time, the singers, at least in certain cases, manifested considerable consistency in intoning different verses during the performance. Most singers' intonation of the three longer notes appearing at the beginning and the end of the excerpt (only those notes were investigated more closely) only differed on average six cents (standard deviation was five cents) across the three verses analyzed. On the other hand, the pitch of the singers' match of the same notes differed by up to 30 cents from the frequency of the corresponding equally tempered scale step value. A detailed description of the results of various studies concerned with the purity of intonation will be found in article II, pp. 331–333 and in article III, pp. 117–119.

An important reason for pitch deviations in playing music or in singing consists in the fact that we tend to operate with musical intervals as categories. Our perception tends to map the combination of any two given sounds onto a discrete category of a particular musical interval, akin to the way that the perception of speech sounds occurs. The boundaries of an interval category may lie at ±50 cents or even more from the centre of the interval (Burns & Ward, 1978; Hall & Hess, 1984). Hence, according to modern views, although acoustically pure intervals may have played a part in the historical development of musical scales, the intonation standard accepted in a culture will still be based on interval categories that have been learned as a result of

enculturation. This means that psychophysical clues such as the absence of beating are, as a rule not used to judge the justness of intervals (Burns, 1999). The intonation standard that currently dominates Western musical culture is based on an equally tempered scale that is slightly compressed for smaller intervals and stretched for bigger ones (Burns, ibid.; for a detailed description, see also article II, p. 332).

One of the aims of articles II and III in the thesis at hand was to investigate the ability of young professional singers to sing a short exercise consisting of a series of melodic intervals (minor second, tritone and perfect fifth) without accompaniment, having only had a brief chance to familiarise themselves with the score. The investigation was conceived to provide an answer to the question of whether the tendencies that have been demonstrated in various tasks performed in connection with stimuli generated in a variety of laboratory conditions can also be observed in the case of singing without accompaniment. The second aim of the study was to use the singers' performance of the exercise as input data for investigating the ability of the singers themselves, as well as of musically trained expert listeners, to identify intonation shifts departing from the corresponding equally tempered values.

The results of the study showed that the differences between individual singers were significant and that some singers' intonations were, in certain cases, outside the conventional categorial range of ±50 cents (see also article II, pp. 336–337 and Figure 4, as well as article III, p. 120 and Figure 2). In performing the exercise, the singers manifested the same tendencies that had been reported in earlier research, in which participants had been asked to tune musical intervals consisting of synthesized sounds. Such tendencies include the proclivity to intone smaller intervals (minor seconds) more narrowly and larger intervals (perfect fifths) more widely than prescribed by the values of the equally tempered scale. They also include a certain hesitation in tuning intervals that are less frequently used in musical compositions (such as the tritone)—this is suggested by the higher standard deviation value.

One of the participants (IO) stood out from the rest, having absolute pitch and appearing to base his judgments primarily on the absolute pitch values of tones instead of the relative interval value between adjacent tones (usual for the other participants). On some occasions this resulted in large deviations in the value of melodic intervals. These exceeded the conventional in-category deviation limit of 50 cents, while the deviations of the respective sounds from the equally tempered scale values intended by the singers were less than 50 cents. Hence, a singer having absolute pitch may, in a performance, produce a series of notes, each within the range of less than 50 cents from the relevant step value of the scale, yet get the required interval category wrong.

We also investigated the ability of two groups of participants to identify pitch deviations from equally tempered values. As previously mentioned, interval sizes have been acquired as part of the general cultural knowledge and are stored in the person's long term memory. We assumed that the participants would be less likely to notice small deviations from the equally tempered values (presumably recorded in their long

term memory) than they would be to identify big deviations. This relationship, however, need not always be linear and may not manifest itself in a uniform manner on a constant basis. The reason for this may lie in the influence of several additional factors such as the tendency to perceive interval sizes categorically, the tendency to prefer the model of a musical scale compressed in the direction of small intervals and stretched for bigger ones, and the tendency to use and to perceive deviating intonations as a means of artistic expression (Sundberg *et al.*, 1995).

In order to simplify our work and because of the limited body of data at our disposal, we decided to disregard the possible impact of such additional factors and to confine ourselves to assessing the functioning of the simple equally tempered model based on the ratio of the fundamental frequencies. The first group, the expert participants, were the singers themselves. It turned out that they were unable, during singing, or immediately after it, to distinguish intervals that had been marked as deviant by the second group (independent listeners) from those that the second group had regarded as intonationally acceptable. The ability of singers to notice intonational deviations in their own performance improved significantly after they had listened to a recording of their performance. In the case of the second group of experts, in order to describe the relationship between the values of intonation deviations and the number of corresponding 'sharp' and 'flat' estimates, we used a linear approximation (see article II, Figure 7); the value of the correlation coefficient was calculated as .59.

Article III basically reproduces the description and results of experiments described in Article II. The only difference is that Article III presents the results of the perception test with the second group (consisting of seventeen independent experts) together with an analysis based on signal detection theory (SDT).

SDT was initially formulated in the 1950s for military applications, in which it was necessary to make probability-based judgments regarding the existence of objects suggested by relatively weak signals appearing against a noisy background and to adjust the sensitivity of the system to a level that was best suited for the particular detection task. SDT is based on the probability, calculated by methods of mathematical statistics, of four possible responses (reactions) to a stimulus: hit, false alarm, correct rejection and miss.

In the 1960s, the method was introduced into perceptual psychology. David Green and John Swets (1966) were among the first of its users. They designed tasks in which the participants were asked to detect weak visual and auditory signals against the background of random noise. In such tasks the result is always shaped by two simultaneous processes. The first is the sensory process, i.e. the process directly related to the activities of the perceptual organs. The second is the decision-making process, i.e. the application of the subjective behavioural strategy chosen by the participant to determine when a signal will be considered as sufficiently strong for it to be reported.

In certain cases there may be an overriding interest not to 'miss' the existence of the signal, for which reason the participant will be susceptible to sound 'false alarms',

that is to report a signal's existence when this is actually not the case. A case in point would be an overzealous student in a solfège class, to whom the teacher has entrusted the task of assessing the purity of the intonation produced by the other students singing an exercise. That student is likely to 'find fault' also where none was committed. Or, vice versa, it may be important to avoid false alarms at any cost, in which case the participant will be likely to ignore a weak signal, reporting an absence of signals even when a relatively strong one is present. This would be exemplified in the actions of a lazy teacher who, in a solfège class, prefers to let the students' small intonation mistakes slip by so as not to make life too hard for him/herself.

The advantage offered by SDT is that it permits the sensory process to be measured separately from the decision-making one, and to focus on a possible bias, and its direction, in decision-making. The sensitivity (*d'*) characterises the size of psychological distance between the signal and background noise, i.e. it provides a measure of how easy it is to notice a signal against the background of noise. If the probability of hits and of correct rejections is an equal 69%, *d'* will equal 1, while in the case of a 75% probability, the value of *d'* will be 1.35.

The implementation of SDT presupposes a clear distinction between the existence of a signal and its absence. It is clearly the case in the task of detecting the presence of a quiet sound. As mentioned previously, it is not as easy to draw a similar line between an interval that is in tune (absence of 'signal') and one that deviates from the standard interval value (presence of 'signal'). For this reason, SDT can only be employed as a heuristic in analysing the behaviour of listeners trying to detect intonation deviations. We can define a certain deviation value, thus designating a boundary between the 'in tune' and 'out of tune' intervals and investigate the sensitivity (*d'*) of experts in relation to that boundary, as well as the tendency of the experts to report the presence (an "out of tune" estimate) or the absence (an "in tune" estimate) of a signal.

In our study, based on the work of Lindgren and Sundberg (1972) on the perception of interval deviations we decided to attribute the value of 20 cents to the boundary. As a method of analysis, SDT permitted us to calculate, for each expert separately, the value of *d'*, i.e. the clarity of that expert's perception of 'signals' (deviations exceeding 20 cents as intonation errors), qualitatively different from 'noise' (deviations less than 20 cents). The value of *d'* for many experts turned out to be less than one. The experts tended to report an absence of signals, that is, they preferred to let pass intervals deviating by more than 20 cents, rather than identify an interval deviating by less than 20 cents as out of tune. We should stress, though, that these results have been obtained in assessments of singing without accompaniment and that although most of the assessors possessed a music education, this was, except for a few, not a university education. Likewise, musical context was all but absent in the exercise.

The last article (IV) of the doctoral thesis reports a study conducted to describe the effect of timbre on perceived pitch in several specifically designed situations that

resemble those encountered in actual music practice. Experiments set up to investigate pitch perception have mostly used a method under which the participants are asked to adjust the frequency of a tone produced by a signal generator to match the pitch of the stimulus (for an example, see Rakowski & Miskiewicz, 1985).

We decided to use a three-alternative forced choice method in which the participants have to assess whether the pitch of the last of two consequently presented tones was either sharp, flat or in tune. Leaving aside the tuning of musical instruments before a musical performance (in tuning, the adjustment of pitch is smooth), this will probably be closer to the way that both musicians and listeners think about intonation during a performance. It stands to reason that the effect of timbre on perceived pitch cannot be markedly significant. If that were the case, the use of electronic tuning aids (most of which measure the frequency of the fundamental) in tuning the different instruments of an orchestra would be pointless, and it would be exceedingly hard for singers to maintain good intonation while producing different vowels (which, in an acoustic sense, could be conceived as sounds characterised by different timbres).

The substance of the study focusses on two different experiments. The first of these involved a group of classically trained singers who were asked to reproduce the pitch of sounds generated as having the timbre of either one or the other of two different musical instruments (the piano and the oboe). The sounds were synthesized on a computer and their pitch varied by intervals of an eighth tone in the pitch regions corresponding to the different locations on the singers' voice range. It turned out that the pitch produced by the singer in matching the stimulus was statistically significantly linked to the timbre of the stimulus, although the extent of the effect was negligible and approximated the value of the pitch difference limen. Altogether, seven professional singers participated in the experiment. The precision of pitch matching (in terms of the frequency of the fundamental) in the case of different singers varied within a range considerably broader than the effect of the timbre of the stimulus.

In the case of the other experiment, expert musicians were asked to assess the pitch produced by one of the singers from the first experiment by comparing it to the pitch of piano-timbre and oboe-timbre sounds having frequencies that were either equal to the fundamental of the singer's pitch or close to it. The experts had to provide an assessment for each pair of sounds, judging whether the singer's pitch was flat, sharp or in tune with that of the instrumental sound.

The singer's pitch was most often assessed as in tune with the instrumental stimulus when the frequency of its fundamental was approximately 20 cents lower than the frequency of the fundamental of the instrumental sounds. The results remained the same also when digital manipulation was used to remove the frequency vibrato from the sound produced by the singer. The probable cause of the phenomenon lies in the difference of the spectral energy distribution of the sounds compared. This is suggested by the fact that sounds whose energy peaked in the region of higher frequencies were perceived as slightly higher.

The pitch shift described in the reported study can be compared to analogous phenomena in other fields of perception. Thus, for instance, already at the end of the 19th century, Franz Müller-Lyer described a visual illusion (Rock, 2004), in which two parallel arrows that have the same objective length and are shown next to one another will be perceived as different in length when, in the case of one of the arrows, the arrowheads point inwards and in the case of the other, outwards. The arrowheads may be regarded as a context the difference in which influences the perception of the principal property of the object (ibid.).

We can suppose that timbre provides a similar context in the case of pitch perception and that it is difficult for the listener to keep sensory information regarding the frequency of the fundamental of the sound completely separate from the information that determines perceived timbre.

To assess the significance of the effect of timbre on pitch perception in the practice of music we should consider pitch deviations having a magnitude of approximately 20 cents in a context that includes information resulting from various studies regarding the pitch difference limen and the deviations of pitch that are observed in live musical performance from the theoretical values of the relevant tuning system. Since the typical pitch difference limen in the case of consecutively presented stimuli in laboratory conditions has been shown to amount to approximately 5 cents (see above), the order of magnitude of 20 cents should be regarded as perceptually significant and capable of influencing the quality of musical performance.

If, however, one takes into account the tendency of the listeners to perceive intervals categorically within a range of approximately ±50 cents, the effect of timbre on the perception of pitch may be regarded as not having a particular significance. Hence we can claim that the importance of the corresponding effect in a real performance situation depends on the specific perceptual context.

Although the pitch shift described above can be approached as an illusion, this should not be regarded as a reason to ignore the phenomenon in real performance situations. Listeners do not normally carry devices for measuring the objective parameters of a musical performance—instead, they come to enjoy the subjective experience. Hence, more studies should be conducted to permit a broader understanding of the phenomena involved. Such studies should include, in addition to live singers as well as piano-timbre and oboe-timbre sounds other stimuli permitting us to compare of the pitches of other more important musical instruments by means of a perception test. The existence of the phenomenon should be investigated for a variety of different pitches, also including the sound level as a third variable.

It is not to be excluded that the magnitude of the pitch shift is also influenced by the existence, and duration, of a pause separating the two stimuli. The perception tests conducted by the authors of the article included a silence of about 2.5 seconds between the stimuli. Such an intermission is likely to spur the listener to create a mental image of the second tone with a pitch that is a perfect match to the pitch of the first sound. This suggests an involvement of the participants' short-term memory rather

than just the echoic one. The echoic memory, as a rule, is relatively short and rarely endures for more than one second (Snyder, 2000). The working memory, the conscious part of which is also referred to as the short term memory (Snyder, ibid.) involves a variety of processes operating on different levels of consciousness. The involvement of the participants' short term as opposed to echoic memory was suggested by certain participants' comments to the effect that in order to perform the test task, they had to make a conscious effort to remember the pitch of the stimulus. To do that, they had to loop it in their memory during the pause that separated the two sounds.

Human echoic memory is supplied with data by our sensory organs, which register the stimuli perceived as a continuous stream of sensory information. On the next level, that of perceptual categorisation, a considerable part of the information registered in the echoic memory will be reduced to a limited number of discrete categories, which will activate the conceptual categories of the long term memory. Hence, it could conceivably be that the pitch shift investigated by the authors occurs during the reduction process referred to above, and would be absent if the silence that separated the two sounds would have been negligibly brief.

2.2.1. The significance of the results of the candidate's pitch-related research

Articles II and III demonstrate that the model of a scale in which minor intervals are slightly compressed, and major intervals stretched in comparison with the equally tempered values, appears to have a universal character. Such a model seems to be valid not only for tuning the intervals of synthesized sounds in laboratory conditions, but also for singing without accompaniment. My research also showed that the deviations (of a magnitude of 20 to 25 cents) from equally tempered values of an isolated melodic interval performed by the singing voice will generally not be perceived as intonation deviations even by listeners with a musical training. Article IV demonstrates that in a pitch comparison task involving sounds with different timbre, the timbre difference may influence the participants' assessment of the pitch of those sounds. The influence may result in deviations whose magnitude is several times higher than the pitch difference limen, so that sounds with a spectral centre of gravity located in the region of higher frequencies are likely to be perceived as having a higher pitch.

3. Conclusions

This doctoral thesis has been inspired by problems and issues that I have encountered in my professional activities as a singer and a singing teacher. The results of the research conducted suggest concrete answers to several specific problems but also lend themselves to wider generalisation. The following conclusions might be highlighted as specifically useful for professional musicians, and for singing training.

As demonstrated by the evidence presented in article I, certain terms which have developed in the course of the history of vocal pedagogy may lack a clearly delimited meaning. Different versions of meaning may 'attach' to the respective term over time, in a series of particular professional or educational situations or simply in the process of communication between the people concerned. The location of the focus of these versions of meaning need not coincide for all of their users and may also differ according to the context of use.

The research results presented in articles II, III and IV suggest the conclusion that, both in performing music or more narrowly in singing, good intonation cannot be universally defined on the basis of the fundamental frequencies corresponding to equally tempered values and the relations between these. In fact, it would be more appropriate to approach such intonation as a compromise between various eventually conflicting tendencies. Thus, even an intonation that is subjectively perceived as 'pure' by a person possessing a musical training can deviate by a magnitude of 20 cents or even more from the value of an equally tempered interval calculated on the basis of the relevant fundamentals.

Similarly, the use of electronic tuning aids will not guarantee a tuning solution that corresponds to the solution assessed as 'pure' on the basis of subjective perception. An intonation that appears to be the best in the context of particular intervals, chords or bars of a piece performed *a cappella* may not be sufficient for that piece to be judged well-intonated as a whole.

Relying on the experience that I have gathered while writing this thesis, I can assure the reader that engaging in scientific research in addition to my professional activities as a musician has not only introduced me to new factual knowledge about my professional field but also improved my intuitive perception of it. That improvement is the result of conducting numerous practical experiments, processing their results and digesting meaning from these (even if they have not always taken the form of a research article). For this reason it is my firm conviction that the results of scientific investigations, and the fact of being part of the process of obtaining these will assist musicians in improving their professional skills—as it has done for

several centuries already. I can only hope that the results of the research reported in the doctoral thesis before you will provide a stimulus for further inquiries and be of practical assistance to musicians who have taken the time to peruse it.

4. Häälekvaliteet ja helikõrgus laulmisel: mõningad taju ja moodustusega seotud aspektid

Allan Vurma

Väitekirja moodustavad neli iseseisvat teadusartiklit, kus tutvustatakse kolme läbiviidud uurimuse tulemusi lauluhääle tajumisest, produktsioonist ning nende terminoloogilisest kirjeldamisest. Esimeses töös, kus teostati tajukatseid sünteesitud lauluhäälele sarnaste helistiimulitega, selgus, et lauluõpetuses kasutatav häälekvaliteetide opositsioon „ees/taga" võib olla seotud mitme erineva hääle akustilise omadusega. Katseisikud tajusid hääle „paiknemist eespool" võrreldes kvaliteediga „taga", kui (1) lauldava vokaali esimese ja teise formandi sagedused paiknesid kõrgemal ja/või (2) lauljaformandi sagedus oli suurem ja/või (3) lauljaformandi suhteline helirõhu tase oli kõrgem. Teise uurimuse käigus läbi viidud produktsioonieksperimendis, kus professionaalsetel vokalistidel paluti laulda kolme erinevat meloodilist intervalli (väike sekund, tritoon ja kvint) nii tõusvas kui laskuvas suunas, selgus, et lauljate intonatsiooni iseloomustasid samad seaduspärasused, mida on täheldatud juba varasematel uuringutel laboratooriumikatsetes generaatorihelide häälestamisel: väikseid intervalle (väike sekund) oli kalduvus laulda vastavast võrdtempereeritud väärtusest kitsamalt ja suuri intervalle (kvint) laiemalt ning kõige ebakindlamalt (suurema varieeruvusega) lauldi muusikapraktikas harva esinevat tritooni. Muusikalise haridusega katseisikud kaldusid hindama puhtalt häälestatuks lauldud meloodilisi intervalle ka siis, kui nende kõrvalekalde suurus vastavast võrdtempereeritud väärtusest ulatus 20-st kuni 25 sendini. Kolmanda uurimuse kolme valikvastusega tajukatses, kus katseisikutel tuli võrrelda erineva tämbriga (klassikalise koolitusega tenorihääl, klaver ja oboe) helide kõrgusi omavahel ning hinnata, kas teisena kõlanud heli oli kõrgem, madalam või sama kõrge, selgus, et lauluhäält tajuti *ca* 20 senti kõrgemana kui samasuguse põhitoonisagedusega klaveri- või oboeheli. Samuti siis, kui lauljatel paluti järele laulda oboe- või klaverihelide kõrgusi, oli lauljate produktsioon oboehelide puhul keskmiselt kuus senti kõrgem kui klaverihelide puhul. Helikõrguste tajukõrvalekalde tekitab tõenäoselt helide spektri raskuskeskme erinev asukoht sagedusteljel.

References

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

Boomsliter, P., & Creel, W. (1971). Toward a theory of melody. Paper presented at the Symposium on Musical Perception, Convention of the AAS, December 28, Philadelphia, PA.

Burns, E.M., & 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.

Carterette, E., & Kendall, R. (1999). Comparative music perception and cognition. In D. Deutsh (Ed.), *The Psychology of music* (pp. 725–791). San Diego, CA: Academic Press.

Deutsh, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The Psychology of music* (pp. 349–411). San Diego, CA: Academic Press.

Dyer, J. (2000). The voice in the Middle Ages. In J. Potter (Ed.), *The Cambridge companion to singing* (pp. 165–177). Cambridge: Cambridge University Press.

Fant, G. (1960). *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*. Mouton: The Hague.

Fletcher, H. (1929). *Speech and hearing*. New York, NY: D. van Nostrand Company.

Frances, R. (1987). *The perception of music.* Hillsdale, NJ: Erlbaum.

Galembo, A., & Askenfelt, A. (2003). Quality assessment of musical instruments—effects of multimodality. 5th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM5), Hannover, September 8–13, 2003.

García, M. (1974). *A complete treatise on the art of singing: complete and unabridged / by Manuel García II.*; the editions of 1847 and 1872 collated, edited, and translated by Donald V. Paschke. New York, NY: Da Capo Press.

Goldstein, J.L. (1973). An optimum processor theory for the central information of the pitch of complex tones. *Journal of the Acoustical Society of America, 54(6)*: 1496–1516.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York, NY: John Wiley and Sons.

Hagerman, B., & Sundberg, J. (1980). Fundamental frequency adjustment in barbershop singing. *Journal of Research in Singing*, *4*: 3–17.

Hall, D.E., & Hess, J.T. (1984). Perception of musical interval tuning. *Music Perception, 2*: 166–195.

Kopiez, R. (2003). Intonation of harmonic intervals: Adaptability of expert musicians to equal temperament and just intonation. *Music Perception*, *20(4)*: 383–410.

von Leden, H. (1997). A cultural history of the larynx and voice. In R.T. Sataloff (Ed.), *Professional voice: The science and art of clinical care* (pp. 7–86). San Diego, CA: Singular Publishing.

Lehmann, L. (1981). *Meine Gesangskunst.* Berlin: Bote & Bock.

Lindgren, H., & Sundberg, J. (1972). *Grundfrekvensförlopp och falsksang.* Stockholm: Stockholm University, Institute of Musicology.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits to our capacity for processing information. *Psychological Review*, *63*: 81–96.

Mori, R.M. (1953). *I maestri del bel canto*. Roma: Casa Musicale A. De Santis.

Newton, G. (1984). *Sonority in singing: A historical essey*. New York, NY: Vantage Press.

Patterson, R.D. (1986). Spiral detection of periodicity and the spiral form of musical scales. *Psychology of Music*, *14*: 44–61.

Plack, C.J., & Oxenham, A. (2005). The psychophysics of pitch. In Plack, C.J., Oxenham A. J., Fay, R.R., and Popper, A.N. (Eds.), *Pitch: Neural coding and perception* (pp. 7–55). New York, NY: Springer.

Rakowski, A., & Miskiewicz, A. (1985). Deviations from equal temperament in tuning isolated musical intervals. *Archives of Acoustics, 10*: 95–104.

Rock, I. (2004). Illusion. In *Encyclopedia Americana*. – International ed. I. Grolier Inc.

Roederer, J. (1973). *Introduction to the physics and psychophysics of music*. Berlin and New York, NY: Springer-Verlag.

Rossing, T.D. (1990). *The science of sound*. Reading, MA: Addison-Wesley.

Sell, K. (2005). *The disciplines of vocal pedagogy: Towards an holistic approach*. Aldershot, Hants; Burlington, VT: Ashgate.

Snyder, B. (2000). *Music and memory: An introduction*. Cambridge, MA: Massachusetts Institute of Technology.

Sundberg, J., Prame, E., & Iwarsson, J. (1995). Replicability and accuracy of pitch patterns in professional singers. *Speech Transmission Laboratory-Quarterly Progress and Status Report 36 (2–3)*: 51–62.

Terhardt, E. (1974). Pitch, consonance and harmony. *Journal of the Acoustical Society of America*, *55(5)*: 1061–1069.

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

Ternström, S., & Sundberg, J. (1988). Intonation precision of choir singers. *Journal of the Acoustical Society of America*, *84(1)*: 59–69.

Vennard, W. (1967). *Singing, the mechanism and the technic*. New York, NY: C. Fisher.

Vurma, A. & Ross, J. (2002). Where is a singer's voice if it is 'placed forward.' *Journal of Voice, 16(3)*: 383–391.

Ward, W.D. (1970). Musical perception. In J. Tobias (Ed.), *Foundations of modern auditory theory* (pp. 405–447). New York, NY: Academic Press.
White, E.G. (1909). *Science and singing*. London: Viscont Music
Wood, A. (1961). *The physics of music*. New York, NY: Dower.

I: Vurma, A. & Ross, J. (2003).

The perception of "forward" and "backward placement" of the singing voice.

Logopedics Phoniatrics Vocology, 28: 19–28.

ORIGINAL ARTICLE

Taylor&Francis
healthsciences

The perception of 'forward' and 'backward placement' of the singing voice

Allan Vurma[1] and Jaan Ross[2]

From the [1]Estonian Academy of Music, Rävala 16, 10143 Tallinn, Estonia and [2]University of Tartu, Ülikooli 18, 50090 Tartu, Estonia

Received 26 March 2002. Accepted 29 January 2003.

Logoped Phoniatr Vocol 2003; 28: 19–28

Singing teachers sometimes characterize voice quality in terms of 'forward' and 'backward placement'. In view of traditional knowledge about voice production, it is hard to explain any possible acoustic or articulatory differences between the voices so 'placed'. We have synthesized a number of three-tone melodic excerpts performed by the singing voice. Formant frequencies, and the level and frequency of the singer's formant were varied across the stimuli. Results of a listening test show that the stimuli which were perceived as 'placed forward', correlated not only with higher frequencies of the first and second formants, but also with the higher frequency and level of the singer's formant.

Key words: formant, placement, singing, synthesize, voice quality.

Allan Vurma, Department of Musicology, Estonian Academy of Music, Rävala 16, EE-10143 Tallinn, Estonia. Tel: +372 6675700. Fax: +372 6675800. E-mail: vurma@ema.edu.ee

INTRODUCTION

In the training of professional singers, instructors often use vocabulary which, to a large extent, is not well understood by the general public. The nature of this vocabulary is metaphoric and aimed at reaching a specific vocal quality, which the instructor has set as the goal. The following list presents examples of this type of vocabulary: 'to support the voice', 'to direct the voice into the mask', 'to make the voice fly'. The aforementioned expressions seem to point to images which a singer is expected to create for her or himself during singing. One may think that such images, as well as the process of their creation, trigger certain unconscious mechanisms which help the singer in reaching the goals of the vocal technique (14). It is possible that the use of metaphoric expressions may be justified because during vocalization the singer perceives the sound in two ways – *via* not only the auditory system but also the vibratory sensations in the head, neck and chest (26). Those metaphoric images evidently have little in common with the scientific understanding of human physiology and acoustical phonetics, despite the fact that some singers believe there to be correlation between the two realms of thought. Some of the metaphoric expressions have been used for a hundred years or more in the training of singers. Their longevity leads one to think that the expressions are fully functional, even when there is not much overlap between them and the scientifically correct description of singing production.

Metaphoric descriptions may be employed in singer training instead of scientific ones because of the complexity of the scientific concepts involved. Singers generally lack a strong background in mathematics and science – rather, they tend to be trained in the humanities, and, thus, a significant effort would be required in order for them to understand the mechanics of the vocal mechanism (8). Even current research professionals do not understand all aspects of voice production, and scientific understanding of one's vocal apparatus does not necessarily help a performer sing better due to the lack of appropriate feedback. For example, Hemsley (13) writes: "It seems to me that for basically healthy singers, the anatomy that really matters is not the anatomy studied by the medical profession, but the anatomy of 'how it feels': the 'as if' anatomy".

It is evident that metaphoric directions cannot be very precise. A metaphor is not useful if the singer lacks associations connected with it, or if the singer's associations differ from those of the instructor. It is important to keep in mind that there are different accepted ways of vocal production in singing (*bel*

 ISSN 1401-5439
DOI: 10.1080/14015430310010854

canto, belting, etc.) and that a verbal expression may denote different things in different singing styles. Also, different schools within the same singing style may set different ideals for the produced voice. A certain expression may be used either by an experienced instructor or by a conductor of an amateur choir. In the first case, one may assume that she or he is sufficiently aware of the different aspects linked to the expression. In the second case, it may well be that the conductor is trying to change the sound quality but, at the same time, her or his idea of the target remains rather fuzzy. We would like to conclude that the use of metaphoric expressions in voice training cannot be avoided. If a metaphoric expression 'works' in a pedagogical sense, i.e., it helps to communicate and to achieve the desired sound, the expression has fulfilled its aim. At the same time, the task of science is to build bridges between the acoustical, physiological and pedagogical terminology. Even if the present use of pedagogical metaphors is fully justified from the pragmatic point of view, we still need to understand the objective contents of the metaphors.

Titze (26) describes that a vowel is said to be in 'focus' if the voice is 'placed' correctly. A vocalist's sensation of where the vowel is localized may also be related to the localization of pressure maxima of the standing waves in the vocal tract.

There have been attempts to find a connection between the terminology describing the sound of voice and its objectively measurable acoustic parameters. Bartholomew (1) has associated the brightness of voice of a good opera singer (the high frequency ring in the vocal timbre) with strong partials in the voice spectrum in the frequency range of 2800 Hz. Cleveland (6) describes connection between the type of voice and the frequencies of vowel formants in it. In the eight male singers that he studied, the vowel formants had lower frequencies if the singer was a bass and higher frequencies if the singer was a tenor, while the formant frequencies corresponding to the baritone were of intermediate value. The qualities of the vocal timbre that are connected to different vocal techniques (e.g., covered, open, throaty, pressed, or free) have been described, on the basis of the spectrograms of one singer, by van den Berg and Vennard (2). Sundberg (19, 20) has studied bass singers of dark and light vocal timbre and discovered that in the case of dark timbre the formant frequencies and the level of the singer's formant were lower than in the case of light timbre. Bloothooft and Plomp (3) analysed vowels that were sung using different vocal techniques like, for example, neutral, light, dark, pressed, free, soft, loud, straight, or with great vibrato. Their acoustic findings seem to suggest that with a dark timbre, the singer's pharynx is wide open and the phonation mode is flow, while the phonation is pressed and the position of the larynx is high in the case of pressed singing. In differentiating the vocal timbre, the most informative term was found to be 'vocal sharpness', in which the level of the higher harmonics was greater.

In another study, Bloothooft and Plomp (4) studied the perception of timbre differences in a vowel sung by eight male and seven female singers using judgements on 21 semantic bipolar scales (e.g., light-dark, open-covered, free-pressed, etc.). Broadly, semantic scales clustered into the categories by vocal technique, general voice evaluation, vibrato, voice clarity, and sharpness. Only sharpness turned out to be a verbal attribute of timbre on which most listeners, regardless of their degree of musical training, agreed in their judgments. Sharpness was found to be acoustically related to the slope of the voice spectrum. The second perceptual dimension (colourfulness) indicated the influence of the relative sound level of the singer's formant. The musicians were able to distinguish between several characteristics of the sung phrases; non-musicians used most semantic scales synonymously.

In this work, we investigate the use of a pair of metaphoric expressions, 'the voice placed forward' and 'the voice placed backward'. We have been interested in whether, and in what direction, the objective acoustic characteristics of the voice change if the singer is trying to bring her or his voice 'forward' or 'backward'.

Random observations from singing lessons at the Estonian Academy of Music point to the fact that the instructors often recommend students to direct their voice forward and to avoid placing it backward. Miller (16) has explained the direction of the voice as a subjective notion which is related to vibratory sensations during singing. Emerich *et al.* (9) claim that a backward voice often has excessive muscle tension and an inadequate vocal technique. According to Emerich *et al.*, a backward voice diminishes effective use of resonators, which in turn causes a smaller level of the singer's formant. A backward voice does not carry well across a concert hall, and a singer with such a voice must work hard in order to make her or himself audible.

In phonetics, however, the 'back-forth' opposition is applied to the vowel quality (18). A front vowel is pronounced so that the tongue is arched in the front of the mouth near the teeth. A back vowel is pronounced so that the tongue is arched in the back part of the mouth near the velum.

We have based the design of the present study on our previous work (28) where we investigated the same terminological opposition of the voice placed forward

or backward. In this work we asked 11 male and nine female voice students from the Estonian Academy of Music to sing melodic triads in contrasting ways so that in one case their voice was placed forward and in another case backward. The melodic triads were performed using five different vowels: /a/, /e/, /i/, /o/, and /u/. Sixteen experts were asked to guess blindly which of the triads were intended to sound with a voice placed forward and which ones with a voice placed backward. Correct guesses were subjected to the acoustic analysis. The results demonstrated that the voice placed forward correlated with the higher sound level of the singer's formant as well as with the higher frequencies of F2 and F3.

In the present study we have extended the task in the following manner. First, synthesized voice stimuli have been used instead of natural voices. Second, some extra parameters (the F1 frequency, for example, as a highly relevant parameter in the perception of vowel quality) have been included in the study, in addition to those of the first study (singer's formant level and the F2 and F3 frequencies). It is rather complicated to determine the F1 frequency in real singing voices, while it is quite easy to manipulate it in the synthesis. Third, we have synthesized and used as stimuli in the experiment some items which were deliberately designed to resemble a voice with the speaking quality rather than with the singing quality. We did it because it is common that singers with insufficient training tend to use this type of voice, as they have not acquired the skills to cluster the upper spectral peaks in order to form the singer's formant. Also, choral singers do not always use the singer's formant technique (24). In order not to overload the experts with too excessive a task we restricted ourselves in the present study to the vowel /a/ only. This vowel is often a preferred one in vocal exercises, and its position among other vowels is sometimes understood as central or neutral (5).

A number of hypotheses were formulated at the beginning of the present study. They are based on our previous work (28) as well as on the long experience of the first author as a professional singer and a singing pedagogue. We proposed that a voice placed forward differs from a voice placed backward in the following ways: 1) the level of harmonics is higher in the region of the singer's formant, 2) the frequency of F2 is higher, and 3) the frequency of the singer's formant is higher. In addition, we wanted to find out whether any systematic differences in F1 frequency can be detected between the voices placed forward and backward. In order to check the above hypotheses we conducted a perception experiment where synthesized items were used as the stimuli. We expected sounds that met the above criteria to be classified by the experts as a voice placed forward.

METHOD

If our aim is to reduce the difference between the synthesized voice stimuli and the stimuli produced by a real singer (or at least to be as close to the desired situation as possible), it is necessary to approximate the synthesized stimuli to a real voice. Sundberg has claimed (23) that a synthesized voice does not sound natural if such parameters have been used in synthesis which are impossible to produce with a real voice. Titze has noticed (26) that anyone who has attempted to mimic the human voice with a computer has been amazed by the unnaturalness that results when the proper variations F0 and intensity are not included in the simulation. The F0 variations in singing are due to vibrato as well as the more general F0 instability caused by the human physiological processes. Minor random F0 variations are always characteristic of a real voice. They do not depend on how much extensive training a singer has been received (17). Therefore, if one wants the synthesized stimuli to meet the criterion of naturalness, it is necessary to apply vibrato and some random variation to the F0.

It is noteworthy that some experts who participated in the present study, claimed that it was impossible to judge voice characteristics on the basis of only one single performed tone (note). They preferred a longer melodic excerpt to be presented, in order to be able to fulfil the task. We have tried to take into consideration this observation in our experimental design, and to make the stimuli to consist of a few tones with different F0, instead of single isolated tones.

In some cases the requirement for naturalness forces us to be less rigorous in design of the synthesized stimuli. Specifically, if we induce a change to a voice parameter (e.g., to the frequency of F*n*), which will result in a highly unnatural sound quality, we then may have to choose to co-vary another voice parameter, together with the F*n* frequency, in order to reduce the net unnaturalness of the sound. A situation like this is not ideal from the point of view of the experimental paradigm, but it occurs frequently in actual voice production where articulatory changes are rarely isolated.

We have synthesized the stimuli using a five-formant source-filter model. This model presupposes that the vocal folds' vibration results in a sound whose spectral envelope decreases linearly at higher frequencies. The source sound, as it passes through the vocal tract, is then modified according to the vocal tract transfer function. This function is described as a number of the formant frequencies which articulatorily are caused by

resonances of the vocal tract (22). The more closely harmonics are situated in the spectrum, the better the envelope of the resulting voice spectrum corresponds to the vocal tract transfer function. This is usually the case with a speaking voice that has a relatively low F0. However, it is also possible to define a formant acoustically, as a maximum in the sound spectrum (10). There is no contradiction between the articulatory and the acoustic definitions of a formant as long as the F0 is low enough. When the F0 increases (which generally is characteristic of female voices) then the frequency distance between the harmonics increases too. This may result in a situation in which two similar spectra may result from different vocal tract transfer functions (23), or in which F0 changes alone may induce significant modifications to the sound spectrum (26).

The frequencies of all formants depend on the configuration of the vocal tract. The frequency of F1 mostly depends on how wide the jaw is open: the wider the jaw, the higher the frequency (22). The frequency of F2 mostly depends on the position of the tongue arch in the mouth cavity: the more frontal the position, the higher the frequency (*ibid*). The main acoustical contribution to the generation of the singer's formant stems from the clustering of F3, F4 and F5 (21). Thus, the sound transfer ability of the vocal tract in the frequency range of these formants increases, and a spectrum envelope peak arises (other things being equal). Experiments with acoustic models of the vocal tract show that such a ring of formants can be attained if the pharynx is wide as compared with the entrance to the larynx tube. It seems that, in many singers, this is obtained by a lowering of the larynx (24). The presence of the singer's formant in the spectrum of a vowel sound is an advantage in that it helps the singer's voice to be heard over an orchestra. The singer's formant technique is used mostly by male voices and low female voices (*ibid*). It is possible, however, to reach a higher sound level at the 3 kHz region by simply singing louder: the level of higher harmonics in this case rises faster than that of the lower ones (22). A singer whose voice classification is higher usually has higher formant frequencies too. The most important factor acoustically in terms of voice classification is the frequency of the singer's formant (23).

STIMULI

Stimuli for the experiment were synthesized by a software called Madde, created by Svante Granqvist at the Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm. Series of melodic major triads were generated with the vowel /a/. Standard acoustic parameters for the synthesis were set on the basis of two voice students from the Estonian Academy of Music, a baritone (B) and a soprano (S), whose voices have earned a good professional reputation. We will call the voices B and S the prototypes. In order for the synthesized voices to sound natural, vibrato was applied to them with the following characteristics: for B, the vibrato rate was set to 5.26 Hz and its extent to 0.52 st, and for S, the rate was 6.25 Hz and the extent 0.5 st. Those values were obtained from the prototype voices. According to Hakes *et al.* (12), the aesthetically optimal vibrato rate amounts from 4.5 to 6.5 Hz and the extent to 0.5 st. These values are in agreement with the values used in the present study. In addition, random F0 variation, of 2 per cent for B and 6 per cent for S, was applied to the stimuli.

In addition to B and S, other stimuli for the experiment were created by modifying a single acoustic parameter in the prototype sounds (see Table 1). In order to determine the amount of variation for each parameter, we tried to follow some guidelines: (a) a change should be large enough to be clearly perceived; (b) the vowel quality should not change from /a/ to some other vowel; and (c) the outcome should remain 'natural', i.e., as similar to the human voice as possible. The following modifications were applied to the prototype stimuli B and S. First, frequency of F1 was increased or decreased (stimuli 2, 7 and 8 in Table 1). Second, frequency of F2 was increased or decreased (stimuli nos. 3, 9 and 10). For the female voice, it was possible to alter F1 and F2 frequencies in both directions without distorting the vowel quality. For the male voice, the formant frequencies could be changed in one direction only: increased for F1 and decreased for F2. Third, increasing the frequencies of the upper formants resulted in the increase of the frequency of the singer's formant. This modification was applied to the male voice only (stimulus 5). The level of the singer's formant was manipulated with an equalizer when necessary, in order to make it equal for the other male voice stimuli used in the experiment. The level variation for the singer's formant did not exceed 1 dB in those cases.

The stimulus B_{spe} was different from the rest of the male voice stimuli because it was designed as to be similar to the speaking voice. For this reason, the F3, F4 and F5 frequencies for the stimulus B_{spe} are separated from each other. For this stimulus, the region which normally corresponds to the singer's formant (i.e., that of F3) was 26 dB weaker than the spectral maximum (see Table 1). In order to increase the naturalness of this stimulus, its vibrato parameters were slightly modified in comparison to the other

Table 1. *Overview of stimuli used in the experiment*

No	Stimulus	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	F5 (Hz)	L_{sf} (dB)
1	**B**	**542**	**1000**	**2520**		**3200**	**−11**
2	$B_{F1\uparrow}$	**732**	1000	2520		3200	−12
3	$B_{F2\downarrow}$	542	**843**	2520		3200	−11
4	B_{spe}	542	1000	**2600**	**3270**	**4500**	**−26**
5	$B_{sf\uparrow}$	542	1000	**2800**		**3400**	−11
6	**S**	**700**	**1150**	**3129**	**3650**	**4030**	**−15**
7	$S_{F1\downarrow}$	**572**	1150	3129	3650	4030	−28
8	$S_{F1\uparrow}$	**900**	1150	3129	3650	4030	−17
9	$S_{F2\downarrow}$	700	**975**	3129	3650	4030	−23
10	$S_{F2\uparrow}$	700	**1345**	3129	3650	4030	−10
11	S_{spe}	700	1150	**3097**	**3939**	**4700**	**−40**

The following acoustic parameters were manipulated in the synthesis: F*n* = the frequency of F*n* (Hz); L_{sf} = the level of the spectral maximum in the singer's formant region (dB); B = male voice stimuli; S = female voice stimuli. The index of a stimulus indicates which of the parameters was modified in that stimulus, with respect to the prototype stimuli B or S: F*n* ↑ = the frequency of F*n* was raised; F*n* ↓ = the frequency of F*n* was lowered; spe = with speech-like timbre and without a singer's formant; sf↑ = with raised frequency of the singer's formant. The prototype stimuli B and S, as well as the specific parameters modified for the individual stimuli, are presented in boldface.

stimuli: the vibrato rate was set to 5.9 Hz and its range to 0.2 st.

The formant frequency values in Table 1 correspond to those actually defined for this synthesis program, except for the male voice stimuli's singer's formant (B, $B_{F1\uparrow}$, $B_{F2\downarrow}$, $B_{sf\uparrow}$) whose frequency values at F3 and F4 in the table were estimated from the LTAS of the stimuli. The level of the singer's formant was also obtained in a similar manner. F5 values, i.e., those for the highest component of the singer's formant, correspond to the actual values used in the synthesis. It is interesting to notice that the frequency of the singer's formant which we have used for the prototype stimulus B (as well as for most of its modifications) corresponds to the singer's formant frequency for a baritone in the data of Dmitriev and Kiselev (7). The frequency of the singer's formant for the stimulus $B_{sf\uparrow}$, however, is close to that of a tenor, in the same data.

There are two possible strategies for synthesis of the singing voice. In the first case we rely on the voice spectrum, and in the second case, on the filter (or vocal tract) characteristics. There is little difference between the two strategies when the stimuli have a low F0, which corresponds with most male voices. For the (female) stimuli with a high F0, however, an attempt to keep the amplitudes of harmonics constant while varying other parameters in the voice, may equal a changing of the vocal tract transfer function of the stimuli. In order to avoid the latter, we abandoned the attempt to stabilize the level of harmonics at the singer's formant region for female voices because the spectral envelope peak might in those cases cover only a single harmonic, which implied that a change of the formant frequency automatically caused a change of level of that harmonic, too. The same phenomenon has been described by Titze (26) when he claims that, as harmonics are being swept through vocal tract formants, large fluctuation in the spectrum may occur.

For the female voice, a higher sound level at about 3.5 kHz in the spectrum is typically observed, which is analogous to the singer's formant of the male voice (22). It does not seem fully justified to call this spectral maximum the singer's formant, however, because its sound level is usually less than that of the male voice. Our approach to the female voice stimuli in this study was similar to the male ones, i.e., we tried to cluster F3, F4 and F5, at the same time keeping other spectral features as close as possible to the prototype stimulus S. This technique seemed to be justified, since it yielded the best result, i.e., a sound quality which resembled the prototype S timbre in the best possible way. An exception was the stimulus S_{spe} which was made similar to the spoken voice and therefore had a greater distance between the upper formants than the other stimuli. The corresponding values of vibrato extent for the stimulus S_{spe} was diminished to 0.15 st and the F0 flutter to 2 per cent.

The Madde synthesis software enables one to produce tones with different F0 by selecting appropriate keys on the simulated piano keyboard which is displayed on the computer screen. The total duration of the synthesized melodic triad was about 5 sec. The sound onset and offset were linearly increasing and decreasing within the time interval of 50 to 100 msec. The individual tones were articulated in *legato*. They were placed in the tonal context of D major for the male voices (F0 being 220, 185 and 147 Hz) and of G major for the female voices (F0 being 587, 494 and 392 Hz). It should be pointed out that the formant frequencies of all stimuli remained within the range

for the Estonian vowel /a/ as measured from samples of normal speech (15).

EXPERTS AND PROCEDURE

There were 22 experts altogether to judge the vocal quality. Among them were 10 professional singers (seven among them were at the same time voice teachers at the Estonian Academy of Music) and eight second- or third-year voice students from the same institution. In addition, four experts (three professional singers and one speech and language pathologist) participated in the experiment *via* the Internet. In the first part of the test, each expert was presented the synthesized melodic triads individually and had to judge on a ten-point scale the extent to which a particular voice stimulus was placed forward. There was a pause after each presented triad, the length of which was determined by the expert. A triad which was performed with a voice placed as forward as possible, was expected to receive the maximum amount (10) of points, and *vice versa*. Experts were also asked to supply written comments on the quality of the voices presented.

In the second part of the test, melodic triads were compared pairwise. In this part, no experts participated *via* the Internet. Their total number was thus equal to 18. The experts had to decide which member of a pair sounded more forward. The experts were given the option of not distinguishing between the members of a pair if they could not detect the difference. There was a three-second silent interval between the members of a pair and a seven-second silent interval between the two successive pairs. High-quality audio apparatus (the *Sony F-345* amplifier, the *Sony CDP-315* CD-player and the *Audes* HiFi loudspeakers) was used on spot, in order to reproduce stimuli to the listeners at the sound level of about 70 dB. We have no information, however, about the audio equipment used by the experts who participated in the experiment *via* the Internet. We believe that top quality audio equipment was not of paramount importance in our experiment, since the stimuli did not contain significant information on frequencies above 5 kHz. Experiments were performed separately for the male and female voice stimuli.

RESULTS

Results of the first part of the experiment are presented in Table 2 and in Fig. 1. In general, experts rated the individual stimuli in accordance with our hypotheses. Voices with a higher frequency of F2 or the singer's formant, were considered to be placed forward, and the stimuli which had a speech-like timbre with no or little singer's formant were considered to be placed backward. Also, the stimuli placed forward have a higher F1 frequency than the stimuli placed backward. According to the results, it seems reasonable to set a boundary between the backward and forward at about six credits on the scale the experts used in their estimation. This boundary is indicated with a dotted line in Fig. 1. The following stimuli would then belong to the voices placed backward: $S_{F1\downarrow}$ with the low F1 value (the average credit 4.5), $S_{F2\downarrow}$ and $B_{F2\downarrow}$ with the low F2 value (the average credits 5.3 and 5.2, respectively), and S_{spe} and B_{spe} which had a speech-like timbre (the average credits 5.5 and 4.6). The following stimuli would belong to the category of placed forward: the prototypes B and S (6.3 and 7.2 credits), the stimuli $B_{F1\uparrow}$ and $S_{F1\uparrow}$ with a high F1 (7.1 and 8.6 credits); the stimulus $S_{F2\uparrow}$ with a high F2 (7.6 credits), and the stimulus $B_{sf\uparrow}$ with the higher frequency of the singer's formant (7.5).

Table 2. *Credits given by experts to the stimuli in their individual assessment*

Stimulus	Average	St. dev.	Max	Min
B	6.3	1.5	10	4
$B_{F1\uparrow}$	7.1	2.6	10	2
$B_{F2\downarrow}$	5.2	2.0	10	2
B_{spe}	4.6	2.2	10	1
$B_{sf\uparrow}$	7.5	1.5	10	4
S	7.2	1.4	9	5
$S_{F1\uparrow}$	8.6	1.4	10	5
$S_{F1\downarrow}$	4.5	1.7	8	2
$S_{F2\uparrow}$	7.6	2.0	10	3
$S_{F2\downarrow}$	5.3	1.8	9	3
S_{spe}	5.5	2.4	10	1

A 10-point scale was used. See also Fig. 1.

Results of the pairwise comparison of the stimuli are presented in Table 3. There is a general agreement between the results in Tables 2 and 3. The higher a stimulus was ranked in Table 2, the more chances it had to be considered more forward placed in the pairwise comparison. If the two stimuli were close in their placement, it was also harder to compare them pairwise – with such pairs, many experts were unable to judge the difference between the two stimuli, or the superiority of one stimulus was apparently decided at random. There are, however, two exceptions. The experts detected a clear difference between the pairs B and $B_{F1\uparrow}$, and B and $B_{sf\uparrow}$ despite the fact that both of the stimuli belong to the group placed forward. These results do not contradict the individual estimates of the stimuli in Table 2. Both $B_{F1\uparrow}$ and $B_{sf\uparrow}$ are ranked higher than B as well as in Table 3 (their scores are 13:4 and 13:1, respectively). We would like to

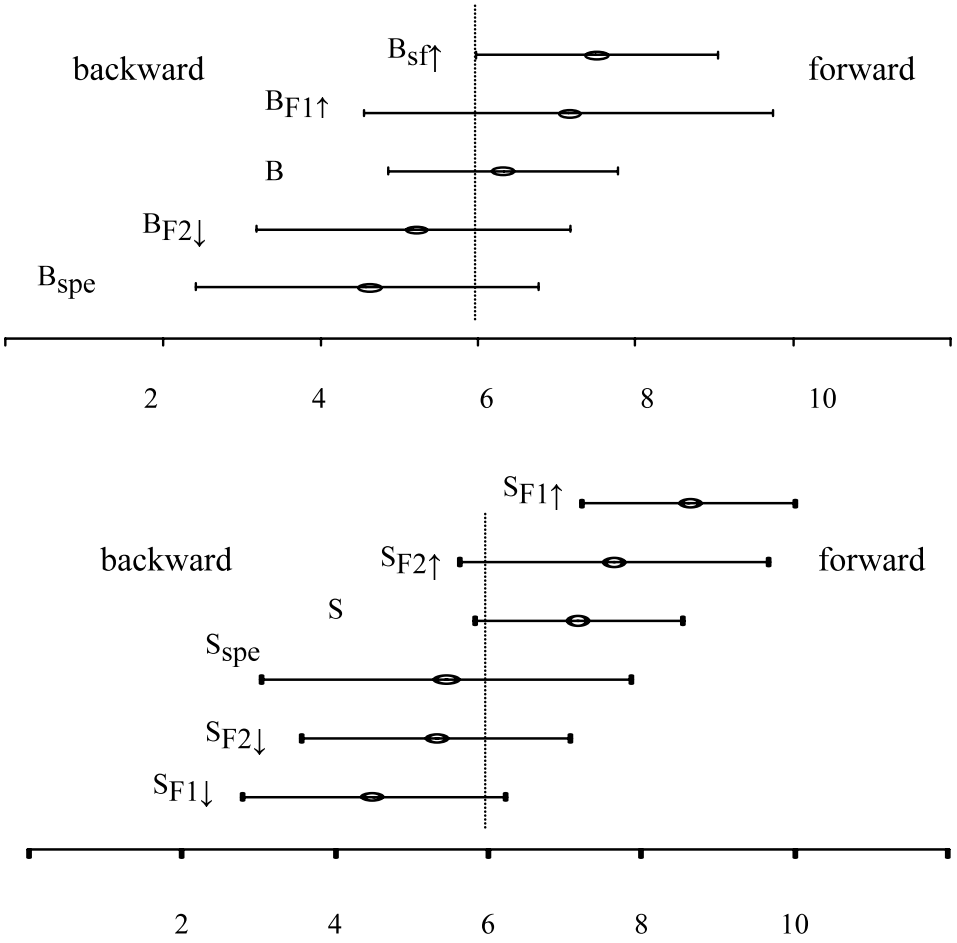

Fig. 1. Average credits given by the experts on a 10-point scale (x-axis) to the stimuli during their individual assessment. The bars represent standard deviation. Data are presented separately for male (above) and female (below) stimuli. See also Table 2.

conclude from this that there was a reasonable agreement between the two experimental conditions (individual estimation or pairwise comparison) in how the experts perceived the backward versus forward placement of the stimuli.

In order to validate the results statistically, we checked the difference of credits given to the stimuli which were classified to the forward and the backward group. Pairwise comparison of the stimuli generally shows a statistically significant difference ($p = 0.05$), as demonstrated by a t-test (Table 3, last column), in credits attributed to each of the two stimuli when they belonged to different categories (forward or backward, see Fig. 1 and Table 2). When the two stimuli belonged to the same category (in other words, when it was hard to distinguish between them), there was no statistically significant difference in credits given to the two stimuli.

Nevertheless, there were some differences between judgments of the individual experts. For example, stimulus B was given credits between four and 10, the average being 6.3. When the frequency of F1 was increased (stimulus $B_{F1\uparrow}$), the average credit to the forward quality increased to 7.1. The credit variance, however, increased, too – the maximum remained 10 but the minimum decreased to two. In the pairwise

Table 3. *Results of pairwise comparison of stimuli in the perception test*

1st stimulus	2nd stimulus	fw and bw	bw and fw	no diff	*F*	df	*p*
B	$B_{F2\downarrow}$	11	4	3	1.86	42	0.04
B	$B_{F1\uparrow}$	4	13	1	–	–	–
B_{spe}	B	4	10	4	2.22(W)	36	0.003
$B_{F2\downarrow}$	B_{spe}	7	5	6	–	–	–
B_{spe}	$B_{F1\uparrow}$	0	14	4	1.42	42	0.001
$B_{F2\downarrow}$	$B_{F1\uparrow}$	2	14	2	1.69	42	0.008
$B_{sf\uparrow}$	B	13	1	4	1.11	42	0.01
S	$S_{F2\downarrow}$	15	1	2	1.65	42	<0.001
S	S_{spe}	10	2	6	3.14(W)	33	0.006
$S_{F2\downarrow}$	S_{spe}	3	7	8	–	–	–
$S_{F2\uparrow}$	$S_{F2\downarrow}$	17	0	1	1.31	42	<0.001
$S_{F1\downarrow}$	$S_{F1\uparrow}$	0	16	2	1.5	42	<0.001
$S_{F2\uparrow}$	$S_{F1\uparrow}$	6	3	9	–	–	–
$S_{F1\downarrow}$	$S_{F2\downarrow}$	1	4	13	–	–	–
$S_{F2\uparrow}$	$S_{F1\downarrow}$	16	0	2	1.38	42	<0.001
$S_{F2\downarrow}$	$S_{F1\uparrow}$	0	17	1	1.58	42	<0.001

The number in the third column (fw and bw) corresponds to the number of experts who estimated the 1st stimulus to sound placed more forward than the 2nd stimulus in a pair. The number in the fourth column (bw and fw) corresponds to the number of experts who estimated the 1st stimulus to sound placed more backward than the 2nd stimulus. The number in the fifth column (no diff) corresponds to the number of experts who did not perceive a difference between the two stimuli. The rightmost three columns present results of a *t*-test which compared whether the difference of credits attributed to the two stimuli in their individual assessment (see Table 2) was statistically significant or not. *F*-values, degrees of freedom (df) and *p*-values are given in the sixth, seventh and eighth columns, respectively. When the variances for the two groups were nearly not the same, the Welch correction (W) was applied. No *F*, df or *p* values are presented for the rows where $p > 0.05$.

comparison of the stimuli B and $B_{F1\uparrow}$ (Table 3), 13 listeners out of 18 found that $B_{F1\uparrow}$ sounded more forward than B while four listeners had an opposite opinion, and one listener could not detect a difference.

It is evident from Table 3 that a comparison of the stimulus with a low frequency of F2 ($B_{F2\downarrow}$) to the stimulus with speech-like timbre (B_{spe}) was complicated for the listeners. Three groups of answers may be distinguished. The first group consists of seven listeners who consider the stimulus $B_{F2\downarrow}$, which has a lower F2 frequency but a strong singer's formant, more forward. The second group of five listeners find that the stimulus B_{spe}, which lacks the singer's formant but has the F1 and F2 frequencies identical to the prototype B, has a more pronounced forward quality. Finally, the third group of six listeners could not tell the difference between the members of the pair. Also, the quality of the female stimulus S_{spe} was not very clear to the listeners. The average credit of its forward quality equalled 5.5, but the credit variance was fairly wide (from one to 10).

In order further to study inter-expert opinion differences we compiled a correlation matrix (Pearson product moment correlation). We ranked the experts according to the strength of correlation of their credit profile with the credit profiles of other experts. We found the number of experts with similar credit profiles (at $p = 0.05$) to vary between zero and 15. Further, we divided the matrix into two groups of 11 experts each, according to the midline. For the first group, the average number of co-experts with a similar credit profile was 11 and the average correlation coefficient between the credit profiles was 0.75. For the second group, these numbers were 3.1 and 0.59, respectively. We interpret the differences between the two groups on the basis of how many student experts belonged to a group – there were seven students in the second group with a low credit profile correlation but only one in the first group with a high credit profile correlation.

Next, we will analyse the verbal comments by listeners given after the experiment. It seems that the forward quality of a voice may mean two different things. First, it may function as a synonym for a sort of ideal voice which is characterized by optimal coordination of the vocal apparatus, according to the standards of the Western classical opera singing. Second, the forward placement of a voice may denote certain specific acoustical qualities of a voice, while the vocal technique at the same time does not need to be optimal. We would like to see the above discrepancy as the main source of variation of the listeners' estimates of voice quality in the experiment. Among the male voice stimuli, B was, according to comments by the experts, frequently considered to possess the best aesthetical and technical quality (some listeners picked up also the stimuli $B_{F2\downarrow}$ and $B_{sf\uparrow}$). The stimulus B_{spe}, on the other hand, was not valued highly and its

quality was estimated as unsupported, passive, dull, relaxed, untrained, or *sotto voce*. The stimulus $B_{F1\uparrow}$ was generally estimated as forward, but at the same time, not of high quality but rather as nasal, narrow, odd, throaty, flat, or with high respiration. From the total of seven female voice stimuli, four (S, $S_{F2\downarrow}$, $S_{F1\downarrow}$, and $S_{F1\uparrow}$) were mentioned as the closest to a listener's ideal standard of voice quality. For the stimulus S_{spe}, it was generally found that the voice was dull, passive or untrained. The stimulus $S_{F2\uparrow}$ was also considered of unpleasant quality (too forward, pressed, nasal, constricted). The majority of listeners did not find it difficult to imagine that the real singers had performed the stimuli.

DISCUSSION

The results of the present study lead one to think that the metaphoric opposition of the voice placed forward versus backward may have different meanings for its users. It seems likely that a request by a teacher to bring the voice forward may point to one or more of the three separate aspects of the voice.

The first aspect is related to the frequencies of F1 and/or F2 which need to be raised. This can be done by opening the mouth wider and/or bringing the tongue arch forward, towards the front teeth. The second aspect is related to a desire to transform the voice with speech-like timbre into one with a clear singer's formant at about 3 kHz region. This goal may be reached with an open pharynx and a relatively low larynx positions. An additional possible benefit derived from the low larynx position is that the distance between the hyoid bone and the thyroid cartilage increases, which in turn decreases the folding of the intraglottal tissue. This means that the vocal folds can vibrate more freely, the voice quality is prone to be better and the mechanical stress on the vocal folds during vibration is lower (27). The third aspect is related to a wish to alter the vocal timbre, in order to make the voice similar to a voice of a different (higher) category (e.g., to change the timbre of a baritone to that of a tenor). It should be noted that a singer has very limited possibilities to accomplish this task because the timbre of each voice category is, first of all, related to the morphology of the vocal apparatus.

Meeting a request to bring the voice forward would therefore require more detailed and longer communication between the singer and the instructor. It is possible to understand the task of placing the voice forward in a fully idiosyncratic way, with little or no overlap of meaning with its use in the wider community. The categories of a voice placed forward or backward may not have a well defined meaning for a student, especially at the beginning of her or his studies. This is demonstrated by the fact that there was little correlation between the expert estimates of voice in the expert group which consisted mostly of students.

Since only the vowel /a/ was used in the stimuli for the present experiment, a question remains to what extent the results obtained are applicable to other vowels. It is apparent that sometimes the relationship between a certain articulatory manoeuvre and the corresponding shift of a formant frequency can be different in the case of different vowels. For example, Sundberg (25) has reported that reduction of the tongue shape bulging lowers the F1 frequency in the case of the vowel /a/, but produces its increase for the vowels /u/, /o/, /e/ and /i/. Further work is therefore needed in order to extend the findings of this study for the vowels other than /a/.

It is important also to recall the characteristics of the human auditory system when discussing the issue of voice placement. One and the same sound level would correspond to different loudness sensations at different frequencies, e.g., 1200 and 5000 Hz (11). One may wish to introduce a category of the centre of gravity of the voice spectrum, which would characterize the sound energy distribution at different frequencies. This imaginary centre of gravity moves towards greater sensitivity of the auditory system as a result of increase in the frequencies of F1, F2 and F3 as well as of increase of the singer's formant level. It seems reasonable to try to connect the concept of bringing the voice forward in the domain of vocal pedagogy, with moving the spectral centre of gravity towards more sensitive regions of the auditory system. The result of this process would be more favourable from the singer's point of view – in order to achieve the same loudness of voice, the required sound level needs to be less and, for this reason, less effort is needed in the vocal production.

According to the theory of speech production by Fant (10), a shift in the frequency of F*n* of a formant, brings about a sound pressure level change of the sound which is mainly confined to frequencies above F*n*. It amounts to plus 12 dB for an increase by an octave of a F*n*. Consequently, if we increase the frequencies of F1 and F2, the sound level in the region of higher formants (including the singer's formant) increases, too.

The character of solving the problems of vocal technique, however, varies for each individual and, to a large extent, depends on the aesthetic ideals set by the singer. We believe, though, that possessing objective information about the operation of the vocal apparatus may enhance the voice development, by clarifying and simplifying the choices a singer must make.

CONCLUSIONS

The terminological opposition of a voice placed forward or backward is widely used in singing instructions. It has several correlations with the acoustic characteristics of a voice, which may occur individually or together. The results of our research have confirmed all the hypotheses presented. A voice placed forward tends to have higher F2 and the singer's formant frequencies and a higher level of the singer's formant. A voice placed forward also has a higher F1 frequency. As a rule, a voice student is recommended to develop the forward quality of her/his voice and to avoid singing with the voice placed backward. However, those categories may have idiosyncratic meaning for various individuals, which, for beginners in particular, makes their use sometimes complicated.

ACKNOWLEDGEMENTS

This work was supported by grant # 3238 from the Estonian Science Foundation.

REFERENCES

1. Bartholomew WT. A physical description of "good" voice quality in male voice. J Acoust Soc Am 1934, 6: 25–33.
2. van den Berg J, Vennard W. Towards an objective vocabulary for voice pedagogy. NATS Bulletin 1959, 15: 10–5.
3. Bloothooft G, Plomp R. Spectral analysis of sung vowels III. J Acoust Soc Am 1986, 79: 852–64.
4. Bloothooft G, Plomp R. The timbre of sung vowels. J Acoust Soc Am 1988, 84: 847–60.
5. Brown OL. Discover Your Voice. San Diego, CA: Singular Publishing Group, Inc, 1996.
6. Cleveland TF. Acoustic properties of voice timbre types and their influence on voice classification. J Acoust Soc Am 1977, 61: 1622–9.
7. Dmitriev L, Kiselev A. Relationship between the formant structure of different types of singing voices and the dimension of supraglottal cavities. Folia Phoniatr 1979, 31: 238–41.
8. Dmitriev L. Osnovy vokaljnoi metodiki (Foundations of the Voice Teaching Methods). Moscow: Muzyka, 2000.
9. Emerich KE, Baroody MM, Caroll LM, Sataloff RT. The singing voice specialist. In: Sataloff RT (ed). Professional Voice: The Science and Art of Clinical Care. New York: Raven Press, 1997 pp. 735–53.
10. Fant G. Acoustic Theory of Speech Production. The Hague: Mouton, 1970.
11. Fletcher H, Munson WA. Loudness: its definition, measurement, and calculation. J Acoust Soc Am 1933, 5: 82–108.
12. Hakes J, Shipp T, Doherty E. Acoustic properties of straight tone, vibrato, trill and trillo. J Voice 1987, 1: 148–56.
13. Hemsley T. Singing and Imagination. Oxford: Oxford University Press, 1998.
14. Jushmanov VI. Vokaljnaja tehnika i ee paradoksy (The Vocal Technique and Its Paradoxes). St. Petersburg: Dean, 2001.
15. Liiv G, Remmel M. On acoustic distinction in the Estonian vowel system. Tallinn: Soviet Finno-Ugric Studies 1970; 6: 7–23.
16. Miller R. On the Art of Singing. New York: Oxford University Press, 1996.
17. Perkell JS, Klatt DH. Invariance and variability in speech process. Hillsdale, NJ: Lawrence Earlboum Associates Publishers, 1986.
18. Report on the 1989 Kiel Convention. International Phonetic Association.
19. Sundberg J. Formant structure and articulation of spoken and sung vowels. Folia Phoniatr 1970, 22: 28–48.
20. Sundberg J. The source spectrum in professional singing. Folia Phoniatr 1973, 25: 71–90.
21. Sundberg J. Articulatory interpretation of the "singing formant". J Acoust Soc Am 1974, 55: 838–44.
22. Sundberg J. The Science of the Singing Voice. DeKalb, IL: Northern Illinois University Press, 1987.
23. Sundberg J. Perceptual aspects of singing. J Voice 1994, 8: 106–22.
24. Sundberg J. Vocal tract resonance. In: Sataloff RT (ed). Professional Voice: The Science and Art of Clinical Care. New York: Raven Press, 1997 pp. 167–84.
25. Sundberg J. My research on the singing voice from a rear-view-mirror perspective. Presentation at the International Conference on the Physiology and Acoustics of Singing, Groningen, the Netherlands, 3–5 October, 2002.
26. Titze IR. Principles of Voice Production. Englewood Cliffs, NJ: Prentice-Hall, 1994.
27. Vilkman E, Karma P. Vertical hyoid bone displacement and fundamental frequency of phonation. Acta Otolaryngol 1989, 108: 142–51.
28. Vurma A, Ross J. Where is a singer's voice, if it is "placed forward"? J Voice 2002, 16: 383–91.

II: Vurma, A. & Ross, J. (2006).

Production and perception of musical intervals.

Music Perception, 23: 331–344.

Production and Perception of Musical Intervals

Allan Vurma
Estonian Academy of Music and Theatre

Jaan Ross
University of Tartu & Estonian Academy of Music and Theatre

This article reports two experiments. In the first experiment, 13 professional singers performed a vocal exercise consisting of three ascending and descending melodic intervals: minor second, tritone, and perfect fifth. Seconds were sung more narrowly but fifths more widely in both directions, as compared to their equally tempered counterparts. In the second experiment, intonation accuracy in performances recorded from the first experiment was evaluated in a listening test. Tritones and fifths were more frequently classified as out of tune than seconds. Good correspondence was found between interval tuning and the listeners' responses. The performers themselves evaluated their performance almost randomly in the immediate post-performance situation but acted comparably to the independent group after listening to their own recording. The data suggest that melodic intervals may be, on an average, 20 to 25 cents out of tune and still be estimated as correctly tuned by expert listeners.

Received July 5, 2003, accepted September 15, 2005

The ability to perform in tune has always been one of the most important professional requirements for a singer, as well as for other instrumentalists who can control pitch on their instrument. Kagen (1950, p. 13) recommends starting with ear training at the very beginning of one's singing studies. According to Kagen, the singer's ability to represent pitches internally and then reproduce them accurately is the main prerequisite for success in professional singing, perhaps even more important than the pleasant timbre of the voice.

Pierce (1999) has claimed that, in Western music, pitch is probably the most essential property of sound. Pitch may be defined as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" (ASA, 1960). Pitch is a subjective quality that cannot be expressed in physical units or measured by physical means. It is, however, strongly related to the repetition rate of the waveform of a sound. For pure tones, pitch may also be influenced by intensity, duration, temporal envelope, partial masking, and (to a certain extent) the ear to which the tone is presented (Moore, 1995). For complex tones, the spectral envelope of the sound is an additional factor that can affect pitch perception. Other parameters, besides the fundamental frequency, usually do not affect pitch perception by more than about ±50 cents (Terhardt, 1988). In the process of perceiving pitch, especially for complex tones, the central nervous system is engaged in addition to the auditory periphery. Different models have been devised in attempts to describe the process in detail (Goldstein, 1973; Terhardt, 1974; Srulovicz & Goldstein, 1983).

In spite of its subjective nature, the pitch of musical tones is still often expressed in terms of the fundamental frequency as the main objective correlate. The fundamental frequency of a tuning fork is usually equal to 440 Hz. This standard for A4, however, varies, sometimes being set about a semitone lower for the performance of Baroque music or slightly higher for a normal symphony orchestra. When intonation in musical performance is measured (e.g., using computer software), the results of such measurements are also most often expressed in terms of fundamental frequency. It has become quite popular in the teaching process to use an inverse procedure as well: to train the intonation accuracy of performing musicians with the help of computer feedback. This method provides a performer with visual fundamental frequency standards that she or he has to match, while the fundamental frequency in performance is measured in real time. The fundamental frequency standard examples in such cases are usually derived from the equally tempered scale.

The role of pitch in music is based primarily on the relationship between two tones (i.e., the ratio of their fundamental frequency, or musical interval) and not on the attributes of single tones (Moore, 1995). In Western

Music Perception volume 23, issue 4, pp. 331-344, issn 0730-7829, electronic issn 1533-8312

music, the most commonly used reference system for pitch relations is equally tempered tuning. In music education and practice, such as solfeggio or preparation for a concert performance, singers often use a keyboard instrument for reference, which, in general, is tuned according to equal temperament. However, the ability to sing in tune during the performance, usually referred to as correct intonation, never corresponds to any of the conventionally defined tuning systems. To a large extent, the exact size of performed intervals often depends on the surrounding musical context, as well as the performer's expressive intentions. A commonly used measurement unit for musical intervals is "cent"; 100 cents are equal to one semitone according to the equally tempered scale.

It is therefore impossible to tell the "correct" pitch value for any of the performed musical tones in terms of a physical parameter: Pitch depends on many factors, both objective and subjective, some of which were outlined above. When considering intonation accuracy, this means that there cannot in principle be a single true fundamental frequency value for a tone (or a single true size for a melodic interval) that could be regarded as the standard for correct intonation. Still, it does make sense to describe the relationship between the performed fundamental frequency values and the equally tempered scale, which may be considered as a kind of arbitrary standard (Burns, 1999). It also makes sense to investigate experimentally how listeners with different musical expertise perceive sung musical intervals as correct or incorrect. The results of such studies, if supported by statistical calculations, make it possible to draw conclusions about general regularities of the musical intonation during a performance. Even if those conclusions are not always valid for each individual listener or performer, they can still be regarded as statistical generalizations about intonational behavior during a musical performance.

In laboratory experiments where the task was to determine the magnitudes of intervals in isolation and not in a musical context, it was found that both melodic and harmonic octaves were tuned with a standard deviation of about 10 cents (Terhardt, 1969; Ward, 1954). The sounds used were sine tones. For other intervals, according to Moran and Pratt (1926), the average deviation varied between 13.5 cents for the pure fourth to 22 cents for the tritone. Rakowski and Miskiewicz (1985) investigated the accuracy of tuning for all 12 intervals of the chromatic scale within an octave, using sine and complex tones, in both rising and falling directions, within the frequency range of 250 to 2000 Hz. They found the interquartile deviation values to be between 20 and 45 cents and showed that there was a tendency to reproduce small or consonant intervals more precisely than larger intervals, as well as to recognize them better. Rakowski (1990) obtained similar results in another experiment where he tried to quantify this phenomenon by introducing a parameter called "interval strength," which is supposed to relate to the strength of memory traces for each particular interval. This strength, in turn, may be related to the frequency of occurrence of the interval in Western tonal music. As examples, smaller intervals occur more frequently than larger intervals, and the tritone interval occurs only rarely (Vos & Troost, 1989).

In musical performance, intonation accuracy depends both on the intentions of a performer and on his or her technical ability to achieve those intentions (Burns, 1999). Sometimes intonation accuracy also depends on the musical genre. Barbershop singing has been found to be characterized by an extraordinarily high degree of intonation precision. Hagerman and Sundberg (1980) measured the average intonation error in barbershop singing to be less than 3 cents. In this style, singers usually do not use vibrato; they can tune harmonic intervals and chords by listening to the beats or roughness between the tones. Vibrato is not thought to influence the accuracy of perception of a given pitch, but it prevents the use of beats as a guide in judging the intonation accuracy of a performed interval (Sundberg, 1987).

There seems to be some evidence of systematic intonation deviations from the equally tempered scale in music practice. Intervals of the magnitude of minor thirds, and less, tend to be performed more narrowly than their equally tempered counterparts. And, conversely, intervals of the magnitude of a minor sixth and greater tend to be performed more widely than the equally tempered standards (Burns, 1999). This is particularly true for octaves and for major and minor sevenths. At the same time, the above rule is not necessarily valid for all performers. Green (1937) and Loosen (1993) have found that in solo violin playing, melodic minor seconds were slightly contracted but melodic major seconds were slightly enlarged. Fourths were found to be close to the natural interval in their study. Similar results were obtained by Ross (1984), who found the average minor second to be 84 cents in solo violin playing but the average major second to be wider than their equally tempered counterparts (207 cents). Rosner (1999) has suggested that intonation behavior may depend not only on interval value but also on the frequency region in which the interval is played.

In the present study we addressed the question of whether regularities observed in production and

perception of isolated musical intervals in a laboratory remained valid in a quasi-musical experimental setting, where the intervals were produced by classically trained singers. Those regularities included different tuning strategies for different intervals depending on their size and degree of consonance or dissonance, as well as differences in the overall stability of an interval as expressed, for example, by the standard deviation in a set of individually tuned intervals. Three intervals were chosen for the study: minor second, tritone, and perfect fifth. The minor second is a small dissonant interval and its size is traditionally regarded as sensitive to the context in music performance studies. The tritone is a medium-sized dissonant interval which is considered to be among the least stable and also regarded as the most complicated to perform in singing and on instruments with free tuning. Finally, the perfect fifth is a medium-sized (or a large) consonant interval regarded as stable in music practice.

The experiment consisted of two parts. In the first part, a group of professional singers was asked to perform a series of melodic intervals. In the second part, listeners, including both independent participants and the performers themselves, were asked to listen to the performance and to estimate whether each interval was tuned correctly. We intended to create a quasi-musical setting for the production experiment. By this we mean that we wanted to minimize differences between a rehearsal routine and the task the singer had to perform in our experiment. The set of intervals used was limited to three so that the task would not become too demanding for participants; however, both ascending and descending intervals were incorporated in the experiment. Performers in the production experiment had to participate in the perception experiment twice: immediately after completing the production task, and after they had listened to a recording of their own performance. We hypothesized that singers might have difficulty evaluating intonation during their performance because of the cognitive and motor load involved in production but may find it comparatively easier to evaluate intonation upon hearing a recording of that performance. Also, during vocalization, the auditory system receives sound both by the air conduction pathway and by body conduction (Sundberg, 1987, p. 159).

These two experiments permitted us to compare the judgments of singers and independent listeners with each other, to assign accuracy scores to each interval produced by each singer, and to determine whether the judgments are related to deviations of the intervals from the equally tempered standards. Further, the intonation standards both in production and in perception can be compared to each other for the three different intervals used in the study. Also, performances by individual singers can be compared to determine whether intonation accuracy judgments are related to their deviation of performed intervals from the equally tempered standards.

Experiment 1: Production

Method

Participants were asked to perform three melodic intervals: a minor second, a tritone, and a perfect fifth, both ascending and descending (see Figure 1). We considered a vocal exercise of such length to be optimal for a singer (i.e., not too long or musically too demanding). A different initial pitch was used for each voice category to make the task as convenient as possible for the participants. Lower voices started the exercise from G4 (contraltos and mezzo-sopranos) or from G3 (basses and baritones), higher voices from B4 (sopranos) or from B3 (tenors). All singers were asked to use a tuning fork to determine the pitch of the initial note in the exercise. The participants were permitted to rehearse the task for a short time, if they chose to do so. They could not, however, use any musical instrument during their rehearsal. All performances were recorded on the hard disk of a computer, using the Kay Elemetrics CSL workstation 4400 and the AKG 420 microphone, which was attached to a headset worn by the singer. The distance between the microphone and the singer's mouth was approximately 3 cm. The recordings were accomplished in a small studio (55 m^2) with a moderate reverberation time of 0.7 s.

Participants were asked to perform the task a total of four times. Initially they were instructed to use the vowel /a/ when singing the requested intervals. During the second trial, they were instructed to use the vowel /i/ instead of /a/ for the middle note of each melodic interval triad (see Figure 1). In the third and fourth trials, participants had to switch back to the original condition. The above procedure was motivated by the work of Ternström, Sundberg, and Collden (1983), who conducted an experiment with choir singers. When the singers had to change the vowel /i/ for /a/ at the middle of a sustained note while instructed to maintain pitch of the note, the fundamental frequency of the sustained note tended to decrease for the portion sung on the vowel /a/.

Participants

Thirteen singers participated in the experiment. All had been studying singing at the Estonian Academy of Music

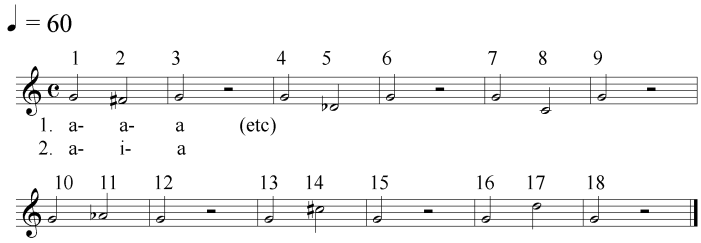

FIG. 1. The singing task in Experiment 1. Each singer repeated the exercise, consisting of 18 notes, four times. Three intervals, minor second, tritone, and perfect fifth, are represented as melodic triads. In the second repetition, the performer was asked to produce the middle note using the vowel /i/ instead of /a/ as in the rest of the trials. Lower voices started the exercise either at G4 (contraltos and mezzo-sopranos) or an octave below at G3 (basses and baritones). Higher voices started the series at B4 (sopranos) or at B3 (tenors).

for at least 4 years. However, they differed to a considerable extent with regard to their ability to sight-read the exercise. Data from four participants were excluded from further analysis because those singers were not able to fulfill the task at an appropriate level. Two of them were not able to sight-read the melodic exercise without a large number of mistakes. A third singer (PR) started the exercise from A4 (instead of G4) in all four trials. The fourth participant (JL) deviated from the starting pitch quite significantly during the course of the experiment, the difference between the beginning and the end of the exercise amounting to more than a semitone. The cause of the inadequacy of those singers may be the insufficient length of their ear-training courses and the emphasis on their vocal development rather than on practicing sight-reading abilities. These vocalists could probably have coped with the singing task much better after a brief rehearsal session with the help of a piano. For example, one of the singers (JL), who was excluded from the group, is nevertheless able to perform successfully major roles in the local opera theater.

Some of the remaining nine participants have had excellent music training and experience. For example, singer MY, besides his M.A. studies at the Music Academy, has been working for several years as a choirmaster at the fully professional, internationally recognized Estonian Philharmonic Chamber Choir (EPCC) and is performing often as a soloist in another smaller vocal ensemble; RJ is a conductor of a semiprofessional, internationally recognized amateur choir and is a frequent soloist as a countertenor in international projects; IO, KE, and RV are singers of the EPCC, IO also frequently performs solos in oratorios and chamber music projects, and RV performs major roles in various chamber opera projects. AK has had her premiere as a soloist in the local opera house. JT has studied singing for 2 years at the Moscow Gnessins Musical Institute and continues her vocal studies at the Music Academy in Karlsruhe, TK is a member of the opera chorus and KT has performed roles in student opera productions and recitals.

The singers were 22 to 30 years old and reported normal hearing. One participant (IO) possessed absolute pitch. Although AP possessors may use a different strategy than relative pitch possessors in the process of intonation, we decided to include IO in the experiment to make possible further comparisons between her and the rest of the group.

Measurement

The recordings of performances were analyzed acoustically with Praat4 software. Fundamental frequency measurement in vocal performance (as well as in the performance of an instrument without fixed tuning) is not a straightforward procedure because fundamental frequency does not usually remain stable during the same note. In real singing, the frequency of a sung tone fluctuates somewhat due to internal "noises" of the human body, such as irregular internal motion of electrical impulses, fluids, and cells within an organ (Titze, 1994, p. 279). A sung tone is also characterized by vibrato of a smaller or larger frequency, the extent of which may be a few semitones. The frequency of a tone may change because the singer lacks sufficient vocal skills: For example, the beginning and end of a note may behave in an unstable way. Usually, if the listener does not pay particular attention to it, voice vibrato and small instabilities in frequency are not perceived as such but rather as a part of the timbre of the voice (Sundberg, 1987). Several studies (e.g., Sundberg, 1978; Shonle & Horan, 1980) have reached the consensus that the pitch of a tone with vibrato is perceived at the average fundamental frequency.

Therefore it seems justified to use the average fundamental frequency value as a representative for the whole note. A stationary part of the sound should be used to determine the fundamental frequency for a particular note. In the present study, the average fundamental frequency was estimated for each performed note and expressed in hertz as well as in absolute semitones relative to 100 Hz (this value is a formal reference used in the Praat software). We tried to estimate the fundamental frequency from a stable segment of sound with a duration of no less than a second, in order to ensure that it was representative for the whole duration of the note. Next, we calculated all interval magnitudes in cents in order to be able to compare them to their equally tempered counterparts. We also calculated the distance of each note in cents from the standard A4 sound produced by the tuning fork that was used in the experiment. The fundamental frequency of the tuning fork was equal to 440.4 Hz at normal room temperature.

Pitches in the recorded performances were estimated using the Praat software. A pilot experiment on pitch matching was conducted to test the reliability of these estimates. Two participants, the authors of this article, took part in the pilot experiment. They had to match a computer-generated complex sound to each note performed and subsequently recorded during the main experiment, listening alternately to a note from the vocal exercise and a computer-generated complex tone. The participants had to tune the pitch of the computer-generated tone to the pitch of the sung note. The computer-synthesized sounds were produced using the Madde software, developed by Svante Granqvist from the Department of Speech, Music, and Hearing, Royal Institute of Technology in Stockholm. All Madde-synthesized complex tones were harmonic and uniform except for the pitch; the spectral envelope of the "voice source" decreased monotonically by 9 dB per octave. The overall level of the signal was equal to about 60 dB.

We preferred to use a complex sound as a matching signal rather than a pure tone because the pitch might have been perceived in a different manner (Greer, 1970) and also because pitch discrimination is better for complex than for pure tones (Spiegel & Watson, 1984). In many cases, the pitch of the harmonic complex tone turns out to be slightly lower than that of the pure tone (Terhardt, 1998). The experts used earphones in their task. The differences between the frequency produced by the computer generator and the average frequency of a sung tone as measured with Praat software were calculated after the end of the pilot experiment.

Results

Pilot experiment. The pilot experiment was aimed at checking whether the fundamental frequency estimation procedure adopted in the present study was reliable enough. Figure 2 presents the histogram of differences between the fundamental frequency of a performed note as estimated with the Praat software and the fundamental frequency that was used for matching the pitch of the performed note. The total number of individual notes performed by all singers was equal to 198. Since data were pooled for the two participants in the pilot experiment, the histogram represents a total of 396 data points.

There was very good correspondence between the pilot experiment matches and the fundamental frequency estimation with Praat. The average difference between the two was only 0.4 cents (the pitch match average was 0.4 cents lower than the Praat estimation average). The standard deviation of the distribution of mismatches was only 13 cents (Figure 2). The overall distribution can be fairly well approximated by a Gaussian curve (one-sample and two-tailed Kolmogorov-Smirnov test for normality yielded $p = .80$). These results are comparable with those of Platt and Racine (1985). In their study, participants with different levels of experience had to match pitches of two sine or complex tones with each other. The typical absolute discrepancy of those matches fell between 10 and 20 cents.

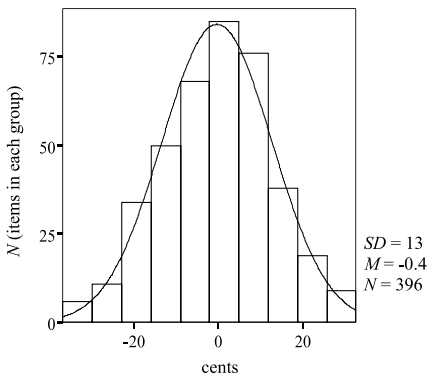

FIG. 2. Results of the pilot experiment. The histogram shows differences between the fundamental frequency of the tone sung by a singer as estimated in the analysis and the fundamental frequency of a synthesizer tone tuned to the same pitch by a participant. The data are pooled for two experts, AV and JR. The continuous line is an approximation to the normal distribution curve.

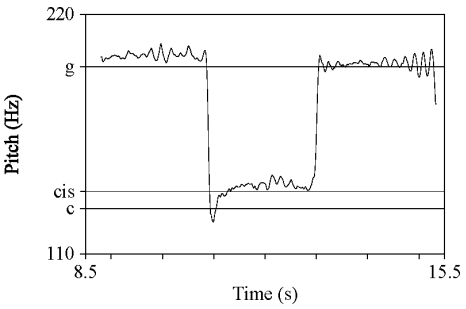

FIG. 3. An example of a performed fundamental frequency curve (singer RV, tritone down). The initial part of the leap extends more than a semitone too low, and only then does the pitch stabilize around its target value.

Only in rare cases was there a remarkable difference between the pilot experiment matches and the fundamental frequency estimation with Praat. It never exceeded 33 cents, however, and those single large mismatches did not concentrate on particular intervals (or performers) but were more or less randomly distributed across the whole data matrix.

A particular case should be pointed out in this connection: an instance when a singer did not "hit" the "correct" pitch level at the very beginning of the note and had to correct it during the note, searching for the proper target. Figure 3 presents a detailed view of a descending tritone as performed by that singer. The beginning of the note was evidently sung a semitone too flat, but the performer was able to correct the mistake in about 0.4 s. If we consider the stable fundamental frequency section of the note as representative of its perceived pitch, then its deviation from the equally tempered standard should be only 6 cents and its status (with a rather large probability) should be estimated as being in tune. This conclusion, however, was not shared by most of the experts who participated in the listening task: 14 listeners out of 17 singled this note out as tuned incorrectly. Also in the pilot experiment, the two participants gave contradictory opinions in this case: The difference between the fundamental frequency match and the Praat-based fundamental frequency estimate was 21 cents for one participant but only 3 cents for the other.

In a few cases, similar problems (i.e., unjustified fundamental frequency instability) were also observed for the initial notes of performed intervals. We estimate that the total proportion of such problematic cases did not exceed a few per cent in the overall data.

Main experiment. The influence of vowel quality (a switch from /a/ to /i/ for the second trial in the series of four) on the performance was checked with a paired *t* test. The test compared the intervals performed in the second trial with the intervals performed in the fourth trial. When we treated all performed intervals individually and pooled the data for the nine participants, we observed a significant effect ($p = .05$) of vowel quality for the descending perfect fifth, which was performed 12 cents narrower with /i/ than with /a/. When we pooled the data for all intervals, we observed a significant effect ($p = .04$) of vowel quality for all ascending intervals, which were performed 6 cents wider with /i/ than with /a/. Because those effects were not very clear and systematic, we avoid more general conclusions as to how a change in vowel quality may influence intonation in singing and limit the use of results from the second experimental trial in further analysis.

Different singers demonstrated different abilities to tune the pitch of the initial note of the interval series (i.e., note #1 in Figure 1). Data on the production of the initial notes for nine participants are presented in Figure 4. Zero on the vertical axis corresponds to the theoretical value of the frequency of the initial note for the corresponding voice category (the equally tempered scale, A4 = 440 Hz). The units on the vertical axis are cents. There are two vertical bars for each of the nine participants depicted in the figure. On each of the left-side bars, the three black squares correspond to fundamental frequencies of the initial notes (#1 in Figure 1) in the first, second, and fourth trials. An empty circle on the same bar corresponds to the fundamental frequency of the initial note in the third trial (this was the trial to be evaluated later during the listening experiment). The right-side (dotted) bar shows variability of the fundamental frequency for all notes with the same pitch (i.e., notes #1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, and 18) in all four trials for each participant. We will call those notes the anchor notes. A diamond corresponds to the average fundamental frequency of the anchor notes in the performance of each participant. The bar extends over the averaged fundamental frequency fluctuation range of the anchor notes (for four trials).

It was fairly common to find deviations from the equally tempered standard of 20 to 30 cents, or even more, for the initial note in the series. The overall distribution of the anchor notes is quite well described by a normal distribution curve ($p = .83$ according to the one-sample Kolmogorov-Smirnov test). The average difference from the equally tempered standard was -4 cents, the median value -3 cents, and the standard deviation 26 cents. Some singers (KT, KE, IO) were quite systematic in placing the initial note (as well as the rest of the anchor notes) higher than the equally tempered

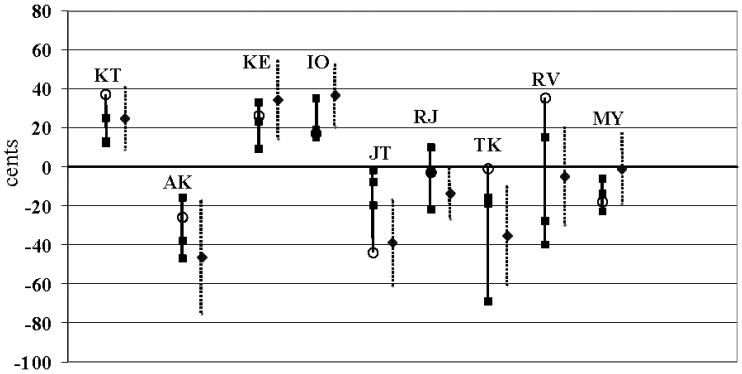

FIG. 4. Analysis of performance in Experiment 1. Data on the intonation of the anchor notes, that is, first and last notes of each melodic interval triad (see Figure 1), are presented for nine participants. The horizontal line at zero corresponds to the expected fundamental frequency of the anchor notes (the equally tempered scale, A4 = 440 Hz). There are two vertical bars for each singer. The left bar connects the initial notes of all four trials; the first, second, and fourth trials are indicated with black squares and the third trial is indicated with an empty circle. The right (dotted) bar corresponds to the extent of variability of the anchor notes, averaged over four trials, and the black diamond to the average fundamental frequency of those notes.

standard, while other singers (AK, JT, TK, MY) consistently did the opposite (i.e., placed it lower).

The pitch of the anchor notes for individual singers could vary by as much as 50 to 55 cents during the production task (e.g., for participants AK and TK, see the length of the dotted vertical bars in Figure 4). For the more stable participants (KT, IO, RJ, MY), the variation was around 30 cents, and the standard deviation (not shown in the figure) was 10 cents or less. It is also possible to track the stability of a singer in the production of the pitch of the anchor notes during the whole task. To accomplish this, one should compare the left-hand (solid) and right-hand (dotted) bars for all singers in Figure 4. The singers KE, IO, and MY apparently showed a tendency to raise the pitch toward the end of the task (the dotted bar is positioned relatively higher than the solid bar), whereas the singers AK, JT, and RJ showed a tendency to lower the pitch (the dotted bar is lower than the solid one).

Some intervals were judged to have been inadequately performed and were removed from further analysis. If an interval was more than 60 cents away from its equally tempered interval counterpart, we considered it possible that the singer may have been confused about which interval she or he was supposed to produce, or confused about how the interval should sound, and omitted it. There were a total of 11 such omissions for the whole group of participants. They occurred more often during the first experimental trial and when the tritone interval was sung.

Using the equally tempered scale as the standard, we compared the intervals measured in the performance of the trials that employed the vowel /a/, that is, the first, third, and fourth (leaving out the data from the second trial, which employed the vowel /i/), with the normative sizes of the equally tempered minor second, tritone, and perfect fifth. Results are presented in Figure 5, in the panel on the left.

Deviations of the performed intervals from their equally tempered counterparts approximate a Gaussian curve. The average difference between the live performance and equal temperament is only about 4 cents. A breakdown of the data into separate interval categories, both ascending and descending, is presented in Table 1. The minor second is generally performed more narrowly as compared to its equally tempered standard, both in ascending and descending modes (the average differences are 11 and 6 cents, respectively). The perfect fifth, on the contrary, is generally performed more widely as compared to its equally tempered standard, also both in ascending and descending modes (the average difference is 14 cents in both cases). The same applies to the tritone (3 and 10 cents sharp for ascending and descending intervals, respectively). Comparison of minor seconds with perfect fifths, as well as minor seconds with tritones, yielded statistically significant differences for both rising and falling intervals ($p < .01$ according to the Mann-Whitney rank sum test). The standard deviation is largest with the tritone (29 cents for the ascending and 24 cents for the descending

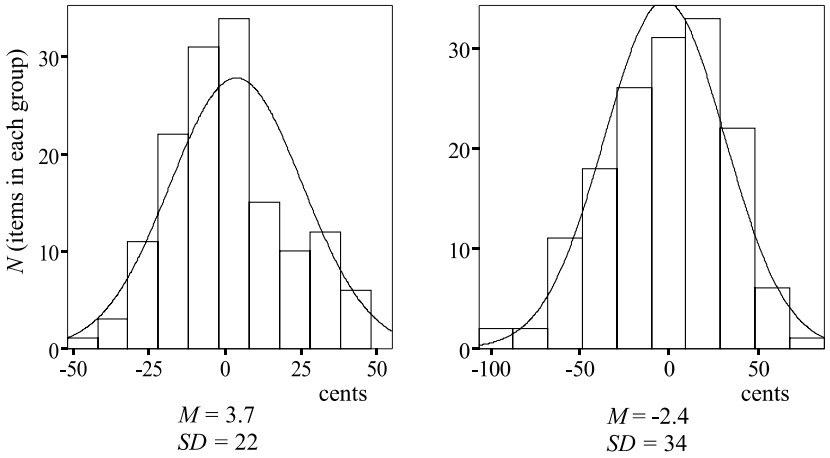

FIG. 5. Deviations from equal temperament in performance (Experiment 1). Left panel: histogram of deviations of the performed intervals from their equally tempered counterparts. Right panel: histogram of deviations of the notes from their counterparts in the equally tempered scale with reference to A4 = 440 Hz. Continuous line: approximation to the normal distribution.

TABLE 1. Deviation (in cents) of the performed melodic intervals from their equally tempered counterparts in experiment 1.

	N	Min	Max	*M*	*SD*
Minor second down	27	−31	11	−6	10
Tritone down	25	−52	50	3	24
Perfect fifth down	26	−18	39	14	16
Minor second up	27	−30	11	−11	11
Tritone up	24	−39	55	10	29
Perfect fifth up	24	−22	51	14	22
All intervals together	152	−52	55	4	22

Note. Data are pooled for nine participants. Also see Figure 5, left panel.

TABLE 2. Deviation (in cents) of the middle notes in the melodic triads, from their standard values on the equally tempered scale.

	N	Min	Max	*M*	*SD*
Minor second down	27	−66	39	3	25
Tritone down	25	−54	64	−2	30
Perfect fifth down	26	−98	41	−17	36
Minor second up	27	−107	50	−12	37
Tritone up	23	−53	87	1	35
Perfect fifth up	24	−69	64	15	33
All intervals together	152	−107	87	−2	34

Note. Calculated with respect to A4 = 440 Hz, in the production task of Experiment 1 (see Figure 1). Data are pooled for nine participants. Also see Figure 5, right panel.

interval) and smallest with the minor second (11 cents for the ascending and 10 cents for the descending interval).

It is possible to analyze the present data in at least two different ways. We can compare the performed intervals individually to their equally tempered counterparts, as we did above, and ignore the fact that a performer might have treated the intervals in the wider context of a scale, the position of which may be fixed on the frequency axis. Another possibility is to fix the whole equally tempered scale relative to A4 = 440 Hz and compare the performed notes to their counterparts in such an ideal scale. For this type of analysis, we calculated the equally tempered comparison intervals on the basis of the fundamental frequency of the tuning fork (440 Hz). In other words, we asked how well a performer was able to hit an equally tempered step built with respect to a reference of 440 Hz as a possible internal standard.

Table 2 and the right-hand panel in Figure 5 display the results of this alternative approach. The two panels in Figure 5 are rather similar to each other. The estimated difference between the equal temperament and the real performance, for all performers and all intervals, turned out to be 4 cents with the first method and -2 cents with the second (also, see Tables 1 and 2). Both distributions are well approximated by a Gaussian curve. The standard deviation was somewhat higher with the second method than with the first (34 versus 22 cents). This implies that the individual intervals must have deviated more from their equally tempered values when analyzed by the second method than when

analyzed with the first method, which is indeed the case; see Tables 1 and 2 for details.

The two methods of analysis may have real relevance in the behavior of singers during an actual performance. The performers may either choose to produce a melody, interval by interval, which would be analogous to the first method of analysis, or try to establish the whole scale internally and use this standard in singing, analogous to the second method of analysis. The intuition of the first author (who is a singer) suggests that, in reality, both strategies may be used in parallel. The difference in standard deviations in our data suggests, however, that the first strategy may have been given priority in the present case. The possibility exists that the internal scale was not sufficiently actualized for a singer because the experimental task consisted of abstract intervals that could not easily have been contextualized musically. It may also be hypothesized that the larger deviations from the scale-step standards resulted from a possible overall shift of the pitch range, which can happen during a performance.

Experiment 2: Perception

Method and Participants

Two groups of participants were used for evaluation of the intervals performed in the first experiment. The first group consisted of the performers themselves. They were tested twice, the first time immediately after the third melodic trial in the production experiment had been completed. Prior to the beginning of this trial, the performers were asked to remember possible intonation errors they might have committed during the performance of this trial and were instructed to mark the incorrectly tuned notes in the score using arrows, an upward-pointing arrow corresponding to a note tuned sharp and a downward-pointing arrow indicating a note tuned flat. The singers completed this task immediately after the end of the trial (or sometimes even during it).

The performers were tested again after they had completed the first, immediate post-performance test after the third trial. For the second test (with the same participants), each participant was played the recording of his or her own just-completed performance and was asked again to record incorrectly tuned notes in the score, this time basing the judgment on listening to the post-performance recording and not on the performance situation.

The second group of participants in the perception experiment consisted of 17 expert listeners. They were asked to judge the intonation of intervals from the third trial of the production experiment performed by the nine singers. The recording was presented for evaluation via a hi-fi amplifying system. The listeners were requested to use the same symbols as the singers themselves (arrows pointing up or down) to indicate those notes which they thought were tuned incorrectly. The recordings were played to them three times, with an approximately 7-s silent interval between the presentations.

Sixteen of the 17 listeners were members of an amateur choir and most of them were musically educated at least at an elementary level. The 17th expert (TO) was a well-known Estonian composer, 45 years old; he has a reputation of possessing a good ear for intonation. This expert also has extensive experience as a vocal group leader as well as a performing singer.

Results for the First Group of Listeners

The singers themselves constituted the first group. They evaluated their own performance twice: the first time immediately after the performance and the second time after they had listened to a recording of their performance. The results for the first, immediate post-performance condition did not reveal any systematic pattern. There were two groups of intervals: the "incorrect" ones, which had been designated as such by the singers themselves, and the remaining intervals, which the singers had accepted as "correct." The average deviations for the two groups of intervals, from their equally tempered counterparts, however, are rather similar: 17 cents for the "incorrect" and 15 cents for the "correct" intervals, the standard deviation in both cases being equal to 15 cents. These differences between the two groups are not significant according to the Mann-Whitney rank sum test ($p = .95$). It seems that judging intervals during the performance itself was not very productive, since the results are close to random.

The results for the first group in the second listening test (when they had to listen to their own recorded performances) were more revealing. The average deviation of "correct" intervals from the equal temperament for all participants and all intervals was equal to 11 cents, and the average deviation of "incorrect" intervals was 25 cents. The standard deviations were 12 and 15 cents, respectively. This difference between the two interval categories is statistically significant ($p < .001$, according to the Mann-Whitney rank sum test).

Results for the Second Group of Listeners

The 17 independent experts estimated 674 (out of 918) intervals as having been performed correctly and 244

intervals performed incorrectly. Figure 6 presents a histogram where all intervals are grouped according to the number of out-of-tune answers attributed to them. Figure 6 shows that there was a significant portion of intervals that all (or nearly all) listeners considered as being performed in tune (left side) and a small number of intervals that were considered as out of tune by the majority of listeners (right side). The distribution is fairly close to normal.

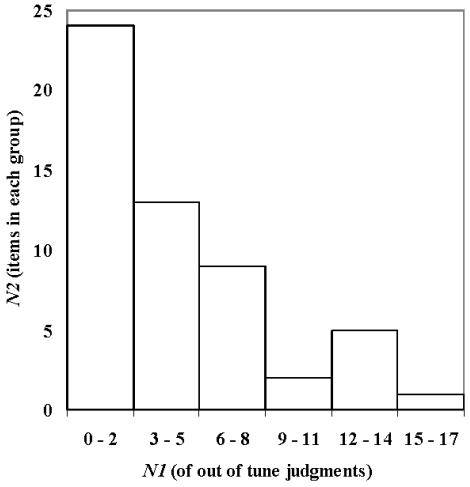

FIG. 6. Distribution of intervals in the perception experiment by the independent group. The intervals are sorted according to the total number (N_1) that were evaluated as out of tune (x-axis). The y-axis represents the number of intervals (N_2) in each group. Only a few intervals received more than 10 out-of-tune judgments (right side), while more than 20 intervals earned only a few out-of-tune judgments (left side). The distribution is close to normal.

In Figure 7, deviations of intervals from their standard values are presented as a function of the number of listeners who estimated an interval as performed out of tune. Deviations from the standard values are calculated with respect to the individual intervals on the (unfixed) equally tempered scale. Points in Figure 7 correspond to the individual intervals. They were approximated by a linear regression curve ($y = 0.13x + 2.25$), with a correlation coefficient value $r(52) = .59$, $p < .0001$ (Pearson Product Moment Correlation). Figure 7 suggests that there is a reasonably good correspondence between tuning of the intervals by singers and their evaluation as in or out of tune by listeners.

In Table 3, the obtained results are grouped according to the interval categories. The columns distinguish between ascending and descending intervals; the number of out-of-tune and in-tune responses is presented in the rows, together with average deviations from the equally tempered standard for each interval category. Table 3 shows that large intervals (tritones and fifths) were more frequently classified as out of tune than seconds. This conclusion holds for both ascending and descending intervals. According to a chi-square test, the differences between small and large intervals are significant for all possible pairwise comparisons: between minor second down and tritone down,

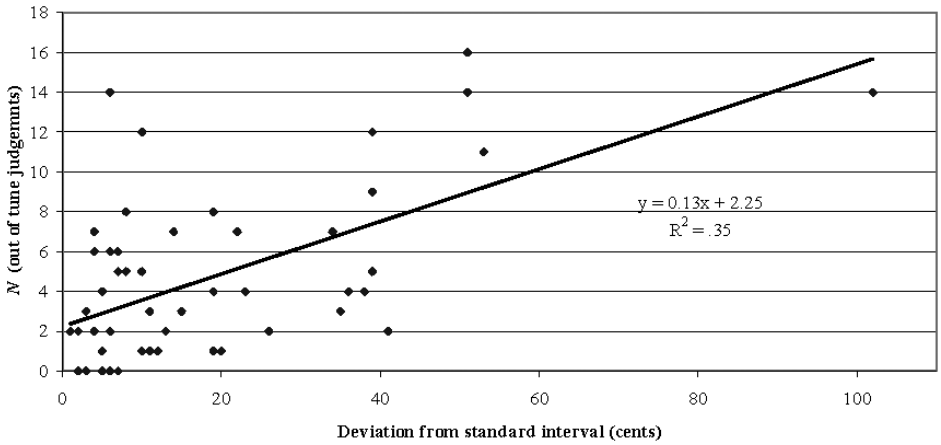

FIG. 7. The number of out-of-tune judgments (y-axis) as a function of interval mistuning (x-axis). Dots represent the intervals. Deviation is calculated with respect to the equally tuned interval sizes.

TABLE 3. Evaluation of performed intervals by a group of independent listeners (N = 17),

	m2 down	m2 up	Tritone down	Tritone up	p5 down	p5 up
# out of tune	6	15	46	48	48	54
# in tune	147	138	107	105	105	99
Dev from interv	6	11	3	10	14	14
SD	10	11	24	29	16	22

Note. The first two rows present the number of out-of-tune and in-tune evaluations, respectively, for all interval categories studied. The average absolute deviation (in cents) from the equally tempered standard value is given in the third row, and the standard deviation is given in the fourth row.

TABLE 4. Intonation accuracy for individual singers.

Singer	MY	AF	KE	IO	JT	TK	KT	RJ	RV
N	13	16	24	26	27	31	34	35	38
M	7	18	8	23	11	23	16	29	22

Note. In the first row, the total number of out-of-tune judgments (*N*) is presented for each singer. The second row gives the average deviation (*M*) of the interval (in cents) from the equally tempered standard value for each singer.

$\chi^2(1) = 35.24$, $p < .001$; between minor second down and perfect fifth down, $\chi^2(1) = 37.8$, $p < .001$; between minor second up and tritone up, $\chi^2(1) = 20.47$, $p < .001$; and between minor second up and perfect fifth up, $\chi^2(1) = 27.02$, $p < .001$.

There is a strong correlation between the standard deviation value (the last row in Table 3) and the number of out-of-tune estimations (the first row) for the interval categories studied: $r(4) = .82$, $p < .05$ (Pearson Product Moment Correlation), which implies that an intonationally more vague interval category tends to attract more out-of-tune judgments from the listeners. As can be seen from Table 3, the largest standard deviation values were obtained for the tritone interval and the smallest values for the minor second, with the perfect fifth falling between the two. Comparison of ascending and descending intervals does not violate this regularity. There is a weak positive correlation, $r(4) = .33$, $p = .52$, between the deviation of the equally tempered standard value (the penultimate row) and the number of out-of-tune estimations for the interval categories in Table 3, but the correlation is not statistically significant.

In Table 4, the results are grouped individually for the nine singers whose data were included in the analysis. The two rows in Table 4 present the number of out-of-tune judgments by listeners and the average deviation from the equally tempered standard for intervals performed by each singer. The correlation between the two rows in the table, $r(7) = .62$, $p = .08$, is statistically significant, but only marginally so. For example, the average deviation for singer MY from the equally tempered standard values was the smallest (7 cents) and he received the lowest number (13) of out-of-tune judgments from the listeners. This singer (MY) was also, intonationally, the most stable in producing the anchor notes, including the initial notes in the four experimental trials, as can be seen in Figure 4.

Finally, we compared the listening task results of the two groups, the singers themselves and the independent listeners. The comparison is somewhat complicated because, as described earlier, the singers evaluated only their own performance, while the independent group evaluated the intonation accuracy in performances of six intervals by nine singers. We divided intervals evaluated by the singers of their own productions into two groups: the out-of-tune intervals and the in-tune intervals. The number of intervals in the first group was 17 and the number of intervals in the second group was 37. Then, we calculated the average number of listeners from the second, independent group who had estimated the same intervals as being out of tune. This calculation resulted in values of 5.9 and 3.9 for the out-of-tune and in-tune intervals, identified as such in the performers' group. The difference between the two groups of intervals was statistically significant, according to the chi-square test, $\chi^2(1) = 13.32$, $p < .001$. This shows that the estimations of intonation accuracy of intervals by the singers themselves, in the post-performance situation, and by independent listeners do not contradict each other.

Discussion

According to Burns (1999), the intonation standard for Western music appears to be a version of the equally tempered scale, which is slightly compressed for small intervals and stretched for wide intervals. A similar tendency was observed in the present study, where singers had to perform isolated intervals. This motivated the adoption of the equally tempered scale as the standard reference for our data. The average difference between the performed intervals and their equally tempered standards was only 4 cents. Ascending and descending minor seconds turned out to be tuned, on average, 11 and 6 cents more narrowly, respectively, as compared to the equally tempered interval; similar average values for the tritone were 10 and 3 cents wider, and for the perfect fifth, 14 cents wider in both directions.

The results confirmed that the singers were able to tune small (seconds) or consonant (fifths) intervals more precisely than tritones. The standard deviation for the minor second was only 10 cents for descending and 11 cents for ascending intervals, but 24 cents for descending and 29 cents for ascending tritones. The standard deviation for perfect fifths fell between the two, being equal to 16 cents for descending and 22 cents for ascending intervals. According to Rakowski (1990), these results may be interpreted in terms of the different "strengths" of intervals, the tritone being "the weakest" of the three. As compared to Rakowski's (1990) data, however, the fifth and second demonstrate the opposite "strength" values in this study, the latter being "stronger" than the former. Vos and Troost (1989) ranked all intervals by their frequency of occurrence in the Western repertoire. If one accepts that "stronger" intervals are those which occur more frequently in music, then one should expect the order of intervals by Vos and Troost (1989) to indicate their "strength." Indeed, the magnitude of the standard deviation for the different intervals in our study is consistent with that order (i.e., minor seconds have smaller standard deviation values than perfect fifths).

Intervals performed by singers may be less stable than intervals produced in a laboratory setting with sine tones, because singing is a much more complex activity than a simple pitch adjustment task. In addition to sensory processes, it also involves a strong motor component. This may explain the slightly higher standard deviation values (on average 22 cents) obtained in our study than, for example, in the study by Burns and Campbell (1994), where the average standard deviation in an adjustment experiment, over all melodic intervals within an octave span, was 18.2 cents.

There was a clear difference between the results of the evaluation task for the first group of listeners, the performers themselves, under two conditions: immediately after completion of the exercise, and after listening to the recording of their own performance once. In the second condition, detection of mistuned intervals was comparable to that of the independent group. The average absolute deviations from equal temperament were 11 cents for the intervals estimated as in tune and 25 cents for those estimated as out of tune (recall that similar values for the independent listeners were 8 and 31 cents, respectively). In the first condition, however, the performing listeners seemed to have operated almost randomly, because the difference in size between the intervals in tune and out of tune was not statistically significant. Obviously the cognitive and motor load on a singer is very high during a performance, while listeners are experiencing music under much less demanding circumstances. This suggests that singers may not be able to adequately judge intonation accuracy during their own performance or immediately following it, without listening to a recording of the performance.

In principle, it is possible to imagine at least two different internal strategies for intonation of intervals in real musical performance. The first strategy would be that in which a singer treats the melodic intervals individually, with little regard for the overall context of musical sounds. This strategy may lead to the accumulation of intonational errors, resulting in a possibly gross deviation from the initial standard of pitch. Another strategy would be to try to maintain an internal standard scale which is always consulted when determining the pitch for the next note to be performed. The present study provides more evidence in support of the first strategy: We obtained a higher standard deviation value when we compared the actual performance to the equal tempered scale fixed with respect to A4 = 440 Hz, than in comparison with the equally tempered intervals whose position was unfixed with respect to the frequency axis (Figure 5). Data in Tables 1 and 2 (cf. the minimum and maximum as well as the standard deviation values) show similar results. The linear regression analysis (Figure 7) also yielded a better fit of the data to the interval model than to the scale model.

There was a participant with absolute pitch (IO) who performed both the singing task of Experiment 1 and the listening task of Experiment 2 (as a member of the first group). Her results were not substantially different from those of the remaining participants. It can be noted, however, that IO seems to have explored the second strategy of intonational behavior, that is, to tune the performed intervals according to an internal scale rather than individually. This is evident from the somewhat larger standard deviation value (34 cents) for the interval-by-interval comparison than for the scale-based comparison (24 cents) of her performance with equal temperament. For all of the participants, those values were on average the opposite, 22 and 34 cents. The overall performance of IO was no more consistent than that of the singers who did not possess absolute pitch. One interval performed by IO (ascending tritone in the first trial of Experiment 1) was removed from further statistical treatment because it was 77 cents narrower than the equally tempered size (but only 41 cents flat of the equally tempered scale step). In general, removal of IO from the group because of her absolute pitch would not have changed the overall results of this study.

Conclusions

In this study, we investigated the production by professional singers and perception by educated listeners of three isolated melodic intervals: minor second, tritone, and perfect fifth. The results demonstrate that the size of produced and perceived intervals can be reasonably well analyzed using equal temperament as a standard reference. Pooling the data for all intervals and nine participants, we determined the average difference between the performed intervals and their equally tempered counterparts to be only 4 cents, with a standard deviation of 22 cents. Our study showed, first, a statistically significant tendency for the diatonic scale to be somewhat stretched, so that smaller intervals (seconds) are performed more narrowly and larger intervals (fifths) more widely than their equally tempered counterparts. Second, smaller (seconds), or more consonant (fifths), intervals were tuned more precisely than tritones, the standard deviation for the former being less than that for the latter. Results suggest that the singers in their performance strategy may have relied more on the individual interval sizes and less on the overall stability of the whole scale. In the listening task, another group, consisting of 17 amateur musicians, had to indicate mistuned intervals in the performance of the previous experiment. Larger intervals (tritones and fifths) were more frequently classified as out of tune than the minor seconds. In addition to the independent listeners, performers from the production experiment had to evaluate their own performance twice, immediately following its completion and after having listened to a recording of the performance. Their evaluations were close to random in the first condition but were comparable to that of the independent group in the second, where the average difference from the equal temperament of the intervals identified as mistuned was equal to 25 cents. Evidence gathered in the two above experiments points to the conclusion that the melodic intervals studied (i.e., minor second, tritone, and perfect fifth) may on average be 20 to 25 cents sharper or flatter with respect to their equally tempered standards and still be estimated as correctly tuned by expert listeners.

Author Note

The authors would like to thank four anonymous reviewers, an anonymous statistician, and the Associate Editor Bill Thompson for their constructive criticism of many aspects of the present article. Comments as well as the language check by Professor Ilse Lehiste and Professor Lawrence Feth of Ohio State University have been very helpful. This work has been supported by research grant #4712 from the Estonian Science Foundation.

Parts of this article were presented during the Fifth International Voice Symposium in Salzburg, Austria, August 2002, the Fifth Triennial ESCOM Conference in Hanover, Germany, September 2003, and the Eighth International Conference on Music Perception and Cognition in Evanston, USA, August 2004.

Address correspondence to: Allan Vurma, Estonian Academy of Music and Theatre, Rävala 16, 10143 Tallinn, Estonia. E-MAIL *vurma@ema.edu.ee*

References

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

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

Burns, E. M., & Campbell, S. L. (1994). Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception? *Journal of the Acoustical Society of America, 96*, 2704–2719.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America, 54*, 1496–1516.

Green, P. C. (1937). Violin intonation. *Journal of the Acoustical Society of America, 9*, 43–44.

Greer, R. D. (1970). The effect of timbre on brass-wind intonation. In E. Gordon (Ed.), *Experimental research in the psychology of music* (pp. 65–94). Iowa City: University of Iowa Press.

Hagerman, B., & Sundberg, J. (1980). Fundamental frequency adjustment in barbershop singing. *Journal of Research in Singing, 4*, 3–17.

Kagen, S. (1950). *On studying singing*. New York: Rinehart.

Loosen, F. (1993). Intonation of solo violin performance with reference to equally tempered, Pythagorean, and just intonations. *Journal of the Acoustical Society of America, 93*, 525–539.

Moore, B. C. J. (1995). *Hearing*. London: Academic Press.

Moran, H., & Pratt, C. C. (1926). Variability of judgments of musical intervals. *Journal of Experimental Psychology, 9*, 492–500.

Pierce, J. R. (1999). The nature of musical sound. In D. Deutsch (Ed.), *The psychology of music* (pp. 1–23). San Diego, CA: Academic Press.

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

Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music, 18*, 60-72.

Rakowski, A., & Miskiewicz, A. (1985). Deviations from equal temperament in tuning isolated musical intervals. *Archives of Acoustics, 10*, 95–104.

Rosner, B. S. (1999). Stretching and compression in the perception of musical intervals. *Music Perception, 17*, 101–114.

Ross, J. (1984). Measurement of melodic intervals in performed music: Some results. In J. Ross (Ed.), *Symposium: Computational models of hearing and vision: Summaries* (pp. 50-52). Tallinn: Estonian SSR Academy of Sciences.

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

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency discrimination tasks by musicians and non-musicians. *Journal of the Acoustical Society of America, 76*, 1690-1695.

Srulovicz, P., & Goldstein, J. L. (1983). A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *Journal of the Acoustical Society of America, 73*, 1266–1276.

Sundberg, J. (1978). Effects of the vibrato and the singing formant on pitch. *Musicologica Slovaca, 6*, 51–69.

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

Terhardt, E. (1969). Oktavspreizung und Tonhöhenverschiebung bei Sinustonen. [Octave enlargement and pitch shift in the case of sine tones]. *Acustica, 22*, 348–351.

Terhardt, E. (1974). Pitch, consonance and harmony. *Journal of the Acoustical Society of America, 55*, 1061–1069.

Terhardt, E. (1988). Intonation of tone scales. *Archives of Acoustics, 13*, 147–156.

Terhardt, E. (1998). *Akustische Kommunikation—Grundlagen mit Hörbeispielen*. [Acoustical communication: Foundations and auditory demonstrations]. Berlin/Heidelberg: Springer.

Ternström, S., Sundberg, J., & Collden, J. (1983). Articulatory perturbation of pitch in singers deprived of auditory feedback. In A. Askenfeldt, S. Felicetti, E. Jansson, & J. Sundberg (Eds.), *Proceedings of Stockholm Music Acoustics Conference SMAC 83, 1* (pp. 291–304). Stockholm: Royal Swedish Academy of Music.

Titze, I. R. (1994). *Principles of voice production*. Needham Heights, MA: Allyn and Bacon.

Vos, P. G., & Troost, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception, 6*, 383–396.

Ward, W. D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America, 26*, 369–380.

III: Вурма А., Росс Я. и Огородникова Е.А. (2006).

Восприятие вокальных музыкальных интервалов.

Сенсорные системы, *20*: 117–125.

СЕНСОРНЫЕ СИСТЕМЫ, 2006, том 20, № 2, с. 117–125

СЛУХОВАЯ СИСТЕМА

УДК 612.85

ВОСПРИЯТИЕ ВОКАЛЬНЫХ МУЗЫКАЛЬНЫХ ИНТЕРВАЛОВ

© 2006 г. А. Вурма, Я. Росс[1], Е. А. Огородникова[2]

Академия Музыки и Театра Эстонии
10143, Таллинн, бульвар Рявала, 16
[1]Тартуский университет
50090, Тарту, ул. Юликооли, 18
[2]Институт физиологии им. И.П. Павлова РАН
199034 Санкт-Петербург, наб. Макарова, 6
E-mail: ogo@infran.ru

Поступила в редакцию 21.10.2005 г.

В работе представлены результаты исследования восприятия и вокального исполнения музыкальных интервалов. Упражнение из трех интервалов (малая секунда, тритон, чистая квинта) в восходящем и нисходящем порядке исполнялось певцами. Восприятие и оценка интонации (чистоты) вокальных интервалов производились двумя группами испытуемых (исполнители-певцы и независимые эксперты).

Показано, что и для вокального исполнения, и для перцептивной оценки интонации характерны редукция малых (секунда) и расширение больших (квинта) интервалов относительно их равномерно-темперированных эталонов. Подтверждена тенденция к более точной настройке и оценке малых и консонантных интервалов (секунды, квинты) по сравнению с диссонирующим тритоном.

Показано, что различия, связанные с частотным диапазоном (начальная нота, голосовая категория), влияют на результаты вокального исполнения в существенно меньшей степени, чем индивидуальные различия. Обнаружено также, что эффективность текущего контроля за интонацией при выполнении певцом вокального упражнения понижена.

В целом, результаты свидетельствовали в пользу категориальной (зонной) природы звуковысотного слуха, а применение теории обнаружения сигналов позволило сделать вывод, что мелодические интервалы с отклонением от эталона в 20–25 центов будут восприниматься как "чистые", т.е. исполненные правильно.

Ключевые слова: восприятие, высота, основной тон, музыкальный интервал, равномерно-темперированный строй.

ВВЕДЕНИЕ

Звуковысотный слух и точность интонирования относятся к важнейшим профессиональным качествам певцов и музыкантов, самостоятельно определяющих высоту тона (настройку) своего инструмента. По мнению Кагена (Kagen, 1950), способность точно воспринимать и воспроизводить высоту звука определяет успех певца даже в большей степени, чем характеристики его голоса. Кроме того, высота звука неразрывно связана с понятием мелодии и относится к основным атрибутам музыки, отражая такую "характеристику слухового восприятия, согласно которой звуки могут быть распределены по музыкальной шкале" (ASA, 1960; Pierce, 1999).

Представляя субъективное качество звука, высота не может быть измерена с помощью только физических единиц. Наиболее адекватным способом ее изучения остается исследование восприятия человека с применением методов психофизики. Известно, что на восприятие высоты влияют различные факторы – интенсивность звука, его длительность, спектр и др. (Moore, 1995). Однако ведущим параметром, определяющим высоту звука, является основная частота (Terhardt, 1988). Разработан целый ряд моделей, отражающих психофизиологические механизмы восприятия высоты звука человеком (Goldstein, 1973; Terhardt, 1974; Szrulovicz, Goldstein, 1983).

Высота музыкальных звуков также традиционно выражается в терминах своего объективного коррелята – основной частоты. Например, основная частота камертона для настройки инструментов (стандарт А4) составляет 440 Гц. Однако ведущую роль в музыке играет соотношение тонов с разной основной частотой или музыкальный интервал (Moore, 1995). Существуют различные системы организации звуковысотных отношений и музыкального звукоряда. Наиболее распространенной из них является равномерная темперация, лежащая в основе настройки целого ряда инструментов. Однако точность воспроизве-

дения высоты (правильная интонация) не определяется только системой настройки. Она зависит от структуры музыкального произведения, а также является одним из выразительных средств исполнителя. Поэтому вариативность интервалов даже при профессиональном исполнении может быть весьма значительной (Rossing, 1990). Сравнение данных игры на скрипке (Ward, 1970) показало, что диапазон вариаций может составлять 78 центов (100 центов = = 1 полутону равномерно-темперированного строя). Средний разброс одноименных интервалов при вокальном исполнении (мелодия из 46 нот) может достигать 104 цента (Frances, 1987). Однако эти отклонения могут не фиксироваться как ошибки интонации даже квалифицированными слушателями.

Таким образом, вариации музыкальных интервалов не всегда обнаруживаются слушателями, хотя известно, что минимальный дифференциальный порог при различении высоты составляет всего 5 центов (Moore, 2000). Эта ситуация может определяться категориальной (зонной) природой звуковысотного слуха. Аналогично восприятию звуков речи, она связана с формированием у слушателя ряда дискретных, культурно обусловленных категорий интервалов, различия внутри которых менее значимы, чем различия между соседними категориями (Гарбузов, 1948). Поэтому большинство слушателей замечает изменение качества интервала только при пересечении условной категориальной границы. В то же время многие из них способны оценить различия и внутри одной категории, когда речь идет о "широте" или "узости" данного интервала (Morrison, Fyk, 2002; Sundberg, 1991; Гарбузов, 1948). Показано, что оставаясь в рамках своей категории, интервал может быть изменен (сжат или расширен) более чем на 50 центов (Burns, Ward, 1978). Например, большая секста, по некоторым данным, переходит в минорную септиму только при расширении, превышающем 80 центов (Hall, Hess, 1984).

Таким образом, оценка чистоты интонации и точности вокальных интервалов представляется трудной задачей. Помимо выделенных факторов этому способствует и отсутствие единых эталонов для измерения значений мелодических интервалов. В этом качестве могут выступать равномерно-темперированные стандарты (Burns, 1999). Сравнение с ними показывает, что в музыкальной практике наблюдаются систематические отклонения интервальных значений. Так, отмечено, что малые интервалы (до малой терции) чаще исполняются более узко, а большие (от малой сексты) – наоборот, более широко (Burns, 1999). Возможны и вариации проявлений этой закономерности. Грин (Green, 1937) и Лузен (Loosen, 1993) показали, что в сольной игре на скрипке малые секунды в основном сокращаются, большие – увеличиваются, а кварты – близки своему эталону. Сходные данные были получены и в другой работе (Ross, 1984). Рознер (Rosner, 1999) предположил, что точность интервала определяется не только категорией, но и частотной областью, в которой он воспроизводится. При вокальном исполнении чистота интонации зависит также от стиля и школы пения (Burns, 1999). В случае пения без вибрато она наиболее высока. Так, средняя погрешность в пении барбершоп (англ. barbershop singing) не превышает трех центов (Hagerman, Sundberg, 1980). Предполагается, что такая точность определяется использованием признаков шероховатости и биений между тонами, недоступных при пении с вибрато (Sundberg, 1987). Кроме того, певцы могут использовать и другие, например, кинестетические ощущения (Mürbe et al., 2004).

Точность оценки интонационной корректности при восприятии музыки также варьирует в широких пределах. Анализ прослушивания вокального материала (10 записей "Ave Maria") показал, что оценки "в тон" характерны для нот с отклонением до 7 центов. Однако и при отклонении в 30 центов они (две ноты в пределах мелодии) могли восприниматься как "чистые" (Sundberg et al., 1996). В другом исследовании (Lindgren, Sundberg, 1972) интонационные ошибки в основном обнаруживались при отклонении свыше 20 центов. В то же время в ряде случаев вариации, превышающие это значение, не выделялись слушателями. Для изолированных интервалов (чистые тоны, сложные сигналы без музыкального контекста) было показано, что стандартное отклонение при настройке октавы составляет 10 центов (Terhardt, 1969; Ward, 1954); чистой кварты – 13.5 и тритона – 22 цента (Moran, Pratt, 1926). По данным Раковского и Мицкевича (Rakowski, Miskiewicz, 1985) интерквартильная широта 12 интервалов в пределах октавы составляет от 20 до 45 центов. Авторами отмечена также тенденция более точного исполнения и оценки малых или консонантных интервалов. Для объяснения этого факта было использовано понятие "прочности интервала" (англ. interval strength), связанное с формированием образов (отпечатков) в памяти слушателя и частотой встречаемости интервалов (Rakowski, 1990).

В настоящей работе было продолжено исследование восприятия и исполнения музыкальных интервалов. Оно проводилось на материале вокальных упражнений, включающих различные по степени встречаемости и трудности воспроизведения интервалы – малая секунда, тритон и чистая квинта. Упражнения выполнялись группой профессиональных певцов с широким диапазоном голосов (от сопрано до баса). Точность интонирования определялась двумя группами слушателей (исполнители-певцы и независимые эксперты).

Предполагалось, что экспериментальное исследование позволит: выявить особенности *вос-*

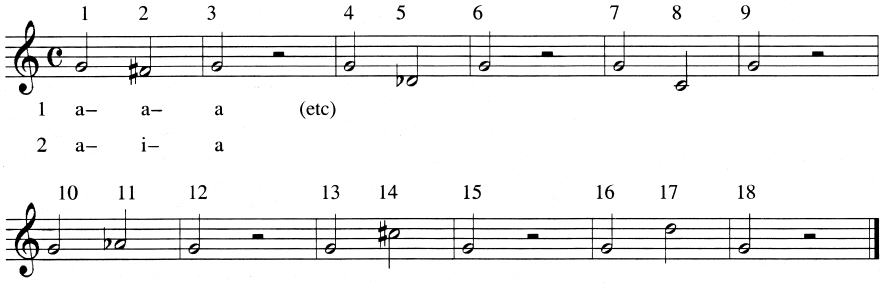

Рис. 1. Пример упражнения в нотной записи (воспроизведение трех мелодических интервалов в нисходящем – ноты 1–9 и восходящем порядке – ноты 10–18).

приятия музыкальных интервалов (категория интервала, частота встречаемости) и *перцептивной оценки* вокальной интонации у разных групп слушателей (исполнители, эксперты); определить *характеристики* интонационной чувствительности слушателей в терминах теории обнаружения сигнала; показать особенности *настройки* и *воспроизведения* интервалов в условиях квазимузыкальной обстановки (минимальный контекст в режиме привычного вокального упражнения) и разного голосового диапазона певцов; оценить эффективность *сенсорного контроля* за интонацией в процессе вокального исполнения.

МЕТОДИКА

Первая часть работы включала выполнение вокальных упражнений, состоящих из трех мелодических интервалов (чистая квинта, малая секунда и тритон) в разном (восходящем и нисходящем) порядке. Пример упражнения в нотной записи приведен на рис. 1. Начальные ноты упражнения соответствовали голосовым категориям: низкие голоса – *соль первой октавы*, 392 Гц (контральто, меццо-сопрано) и *соль малой октавы*, 196 Гц (басы, баритоны); высокие голоса – *си первой октавы*, 494 Гц (сопрано) и *си малой октавы*, 247 Гц (теноры). Настройку проводили при помощи камертона. Упражнение исполнялось 4 раза: 3 раза только с гласным /a/ и 1 раз (второе исполнение) с переходом в середине триады на гласный /i/ (см. рис. 1).

Запись упражнений проходила в студийных условиях (55 м^2; среднее время реверберации 0.7 с) при помощи компьютера, системы Kay Elemetrics CSL4400 и микрофона AKG420. Расстояние между микрофоном и ртом певца составляло 3 см.

В качестве испытуемых-исполнителей выступали 13 певцов в возрасте от 22 до 30 лет, обучавшихся пению не менее 4 лет (Музыкальная академия Эстонии). Все певцы обладали нормальным слухом. У одной испытуемой (исп. IO) был также зафиксирован абсолютный музыкальный слух.

Анализ записей проводили при помощи программы Praat'4. В качестве показателя вокальной интонации выступало среднее значение F0, что определялось известной нестабильностью основной частоты и проявлениями вибрато, свойственными вокальному исполнению (Titze, 1994; Sundberg, 1978; 1987; Shonle, Horan, 1980). Измерения проводили на стабильном сегменте звука (длительность не менее 1 с) и оформляли как в герцах (Гц), так и в абсолютных полутонах относительно 100 Гц (точка отсчета в программе анализа). Значения интервалов оценивались в центах. Для сравнения использовались эталоны равномерной темперации.

Во второй части работы проводили перцептивную оценку данных вокальных упражнений. В качестве испытуемых выступали две группы слушателей – певцы (исполнители) и независимые эксперты. Группа певцов проходила тест дважды: 1-й раз – после выполнения вокального упражнения, 2-й раз – после прослушивания своего исполнения в записи. Испытуемых-певцов просили отмечать обнаруженные интонационные ошибки в нотной записи. Для их обозначения использовались "стрелки" ("вверх" – нота завышена, "вниз" – нота занижена).

Группа экспертов состояла из 17 человек, которые не были профессиональными певцами, но имели музыкальное образование, 16 из них пели в любительском хоре, а один (исп. TO, 45 лет) – был композитором и обладал опытом работы в ансамбле. Эксперты оценивали интервалы на материале записей вокальных упражнений. Инструкцию не меняли – ошибки исполнения отмечали стрелками. Прослушивание проводили 3 раза с перерывом в 7 с.

Для описания данных перцептивной оценки был использован аппарат теории обнаружения сигналов, позволяющей разделить сенсорные и психологические показатели чувствительности

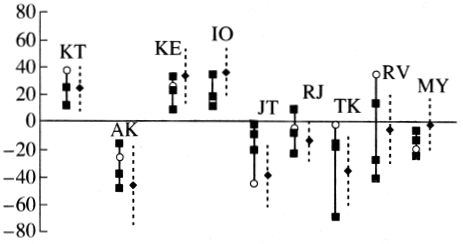

Рис. 2. Результаты исполнения вокального упражнения (данные девяти испытуемых-певцов).

По оси ординат – отклонение F0 начальной ноты от стандарта равномерной темперации (в центах). Обозначения: вертикальная сплошная линия – данные для начальных нот упражнения (черные квадраты – 1-, 2- и 4-ое исполнение, белый круг – 3-е исполнение упражнения, которое использовалось для перцептивной оценки интонации).

Вертикальная точечная линия – разброс ключевых нот и их среднее значение (ромб).

(англ. sensitivity) и установки (англ. bias), связанной с выбором критерия (Macmillan, Creelman, 1991). Ответы испытуемых (экспертов) распределяли по четырем основным категориям: *попадание* (порог отклонения превышен и интервал воспринимается как фальшивый, исполненный "не в тон"); *промах* (отклонение больше порога, но интервал воспринимается как чистый); *ложная тревога* (отклонение от эталона меньше порога, но интервал воспринимается как фальшивый) и *покой* (отклонение меньше порога и интервал воспринимается как чистый, исполненный "правильно").

Пороговый критерий для разделения категорий был выбран на основе данных работы (Lindgren, Sundberg, 1972), отмеченной во введении, и составлял 20 центов.

РЕЗУЛЬТАТЫ

Анализ вокальных упражнений выявил разный уровень профессиональной подготовки испытуемых. Большинство из них (девять певцов) составили однородную по качеству исполнения группу. Это – два сопрано (исп. KT и AK), два меццо-сопрано (исп. KE и IO), один контратенор (исп. RJ), один тенор (исп. MY), один баритон (исп. RV) и один бас (исп. TK). Но четыре певца допустили слишком большое число ошибок при чтении с листа и воспроизведении начальной ноты упражнения. Поэтому записи их вокального исполнения были исключены из дальнейшего рассмотрения.

Сравнение данных при выполнении упражнений показало также, что качество гласного оказывает значимое ($p = 0.04$) влияние на интонацию вокальных интервалов восходящего ряда. Так, интервалы с опорным гласным /i/ исполнялись, в среднем, на 6 центов шире, чем с гласным /a/. Сходный эффект влияния качества гласного на чистоту интонации в хоровом пении был отмечен и в работе (Ternström et al., 1983). Этот результат определил проведение анализа точности вокальной интонации на материале упражнений с использованием только одного гласного – гласного /a/.

Результаты анализа интонации в первую очередь свидетельствовали о значительных различиях в способности певцов настраивать высоту начального тона. На рис. 2 показаны отклонения основной частоты (F0) начальных нот от значения равномерно-темперированного эквивалента (РТЭ). Там же приведены данные (среднее значение и разброс F0) для всех *ключевых* нот (ноты с одинаковой высотой тона – № 1, 3–4, 6–7, 9–10, 12–13, 15–16, 18 на рис. 1). Видно, что отклонения начальной ноты от РТЭ могли достигать 30 и более центов. В целом данные для ключевых нот хорошо соответствуют нормальному распределению ($p = 0.83$, тест Колмогорова–Смирнова): среднее отклонение от эталона – минус 4 цента; стандартное отклонение – 26 центов. Можно отметить также, что ряд певцов (исп. KT, KE, IO) систематически завышали ключевые ноты, другие (исп. AK, JT, TK, MY), наоборот, последовательно занижали их.

На основе результатов вокального исполнения была сделана попытка проверить действие эффекта расширения октавы (Terhardt, 1998). Частота начальной ноты для низких мужских голосов была на октаву ниже низких женских. Проявление эффекта означало, что для низких мужских голосов начальный тон будет ниже, чем расчетная чистая октава. Эта тенденция подтвердилась при сравнении данных меццо-сопрано (исп. KE и IO) и баса (исп. TK). Однако для меццо-сопрано (исп. JT) и баритона (исп. RV) ситуация оказалась обратной (см. рис. 2). В целом, данные показали, что различия, связанные с частотным диапазоном (начальная нота, голосовая категория), влияют на результаты в гораздо меньшей степени, чем индивидуальные различия по группе испытуемых. Так, высота ключевых нот в индивидуальном исполнении могла колебаться в пределах до 50–55 центов (исп. AK и TK). У более стабильных исполнителей (певцы – KT, IO, RJ, MY) разброс высоты составлял 30, а стандартное отклонение – 10 центов.

Результаты сравнения вокальных интервалов с РТЭ приведены на рис. 3, *а*. Видно, что полученное распределение (величина отклонения от РТЭ) близко к нормальному, а среднее различие между вокальными и эталонными интервалами составляет 3.7 цента.

В табл. 1 представлены сводные данные по всем категориям интервалов. Можно отметить, что малая секунда исполняется у́же, чем РТЭ. В среднем различия для восходящего и нисходящего ряда составляют 11 и 6 центов соответственно.

Чистая квинта, напротив, чаще воспроизводится шире (среднее отклонение от РТЭ в обоих направлениях составляет 14 центов). Сходная тенденция наблюдается и для тритона, который в среднем превышает РТЭ на 3 и 10 центов (восходящий и нисходящий ряд соответственно). При сравнении результатов были выявлены статистически значимые различия для восходящего и нисходящего порядка исполнения ($p < 0.01$ по тесту Манна-Уитни). Показано также, что максимальное стандартное отклонение соответствует тритону (с учетом ряда – 29 и 24 цента), а минимальное – малой секунде (11 и 10 центов).

При анализе полученных данных были рассмотрены две процедуры сравнения, отражающие возможные стратегии настройки певца – настройка отдельных интервалов вне контекста музыкальной шкалы и настройка с учетом контекста шкалы как целостной звуковысотной системы с фиксированным эталоном высоты для частоты *ля первой октавы* – 440 Гц.

Результаты анализа с применением обоих подходов – вне контекста и с учетом контекста шкалы (стратегия 1 и 2 соответственно) – приведены в табл. 1, 2 и на рис. 3, *а*, *б*. Видно, что оба распределения близки к кривой Гаусса и сходны между собой (среднее отличие от РТЭ составляет 4 и –2 цента, соответственно). Однако величина стандартного отклонения во втором случае заметно превышает значение, полученное вне контекста частотно закрепленной шкалы – 34 против 22 центов.

Результаты второй части работы были представлены данными по восприятию и перцептивной оценке вокальных музыкальных интервалов как исполнителями (певцами), так и независимыми слушателями (экспертами). Группа испытуемых-певцов проходила этот тест дважды и оценивала корректность только собственного исполнения. Результаты теста № 1 (после выполнения вокального упражнения) включали две категории ответов: "неправильные" интервалы (оценка "не в тон") и "правильные" – все остальные. Средние значения отклонения от РТЭ для них были близкими – 17 и 15 центов (для "неправильных" и "правильных" интервалов соответственно), а стандартные отклонения – одинаковыми (15 центов). Таким образом, различия между категориями ответов оказались не значимы ($p = 0.95$; тест Манна-Уитни). Это позволило заключить, что перцептивные оценки интонации после исполнения упражнения имеют практически случайный характер, а текущий контроль за точностью интонирования при выполнении вокальной задачи у певцов нарушен.

Результаты оценки своего исполнения в записи (тест № 2) были гораздо точнее. Среднее отклонение "правильных" интервалов от равномерной темперации (для всех испытуемых и интервалов) составило 11 центов, среднее отклонение

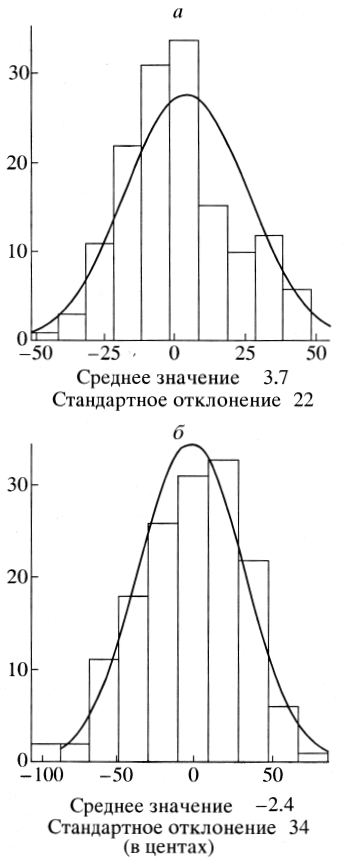

Рис. 3. Отклонения вокальных интервалов от равномерной темперации.

а – распределение, соответствующее стратегии № 1 (расчет для интервалов вне контекста шкалы); *б* – распределение, соответствующее стратегии № 2 (расчет с учетом контекста равномерно-темперированного строя и стандарта A4 = 440 Гц).

"неправильных" интервалов – 25 центов. Стандартные отклонения равнялись 12 и 15 центов, соответственно. Различие между основными классами ответов в этом случае было статистически значимым ($p < 0.001$; тест Манна-Уитни).

При анализе результатов, полученных для группы испытуемых-экспертов, выделялись три категории интервалов: "неправильные" (оценки "не в тон" у большинства испытуемых, $N \geq 8$); "правильные" (оценки "не в тон" не более чем у одного испытуемого; $N \leq 1$); "амбивалентные" (оценки "не в тон" у части испытуемых; $1 < N < 8$). Эти данные приведены в табл. 3. Можно видеть,

Таблица 1. Результаты сравнения исполненных вокальных интервалов с их равномерно-темперированными эквивалентами (значения отклонений, в центах)

Ряд	Интервал	Количество	Минимум	Максимум	Среднее	Стандартное отклонение
Нисходящий	Малая секунда	27	–31	11	–6	10
»	Тритон	25	–52	50	3	24
»	Чистая квинта	26	–18	39	14	16
Восходящий	Малая секунда	27	–30	11	–11	11
»	Тритон	24	–39	55	10	29
»	Чистая квинта	24	–22	51	14	22
Все интервалы		152	–52	55	4	22

Таблица 2. Результаты сравнения высоты средних нот в мелодических триадах с эквивалентами равномерной темперации при фиксации стандарта A4 = 440 Гц (значения отклонений, в центах)

Ряд	Интервал	Количество	Минимум	Максимум	Среднее	Стандартное отклонение
Нисходящий	Малая секунда	27	–66	39	3	25
»	Тритон	25	–54	64	–2	30
»	Чистая квинта	26	–98	41	–17	36
Восходящий	Малая секунда	27	–107	50	–12	37
»	Тритон	23	–53	87	1	35
»	Чистая квинта	24	–69	64	15	33
Все интервалы		152	–107	87	–2	34

Таблица 3. Результаты экспериментов по восприятию вокальных интервалов (данные группы экспертов)

Оценка интервала	Количество	Ответы "не в тон"	Отклонение от эталона (в центах)		
		Число экспертов (среднее по группе)	минимум	максимум	среднее
Неправильный (фальшивый)	10	11.8	–10	52	31
Правильный (чистый)	14	0.57	–20	12	8
Амбивалентный	30	3.9	–39	41	15

что в категорию "неправильных" попало 10 интервалов, в группу "правильных" и "амбивалентных" – 14 и 30 интервалов соответственно. При этом для "неправильных" интервалов средние отклонения от РТЭ составили 31 цент (максимально до 52), для "правильных" интервалов – только 8 центов (максимально до 20). Амбивалентные интервалы заняли промежуточное положение: среднее отклонение от РТЭ – 15 центов (максимальное – до 41). Различие между основными категориями интервалов ("правильные – неправильные") было статистически значимым (p = 0.002; тест Манна-Уитни).

Оказалось также, что при восприятии музыкальных интервалов проявляется тенденция, близкая к закономерности, обнаруженной при исполнении – редукция малых и расширение больших интервалов. Например, две квинты восходящего ряда с отклонением от РТЭ всего –8 и –10 центов стабильно оценивались как фальшивые. В то же время малые секунды на 19 (нисходящий) и 20 (восходящий ряд) центов более узкие, чем РТЭ, получили оценку "в тон" у большинства экспертов (16 из 17).

Результаты восприятия вокальных интервалов испытуемыми второй группы были рассмотрены в рамках теории обнаружения сигналов. Это позволило описать характеристики интонационной чувствительности слушателей в соответствующих терминах и оценить установленный пороговый критерий. Полученные данные представлены на рис. 4. Отдельные точки соответствуют экспертам, чувствительность которых выражается показателем d' (обнаружимость сигнала). Видно, что практически все точки лежат выше диагонали и $d' \geq 0$. Это означает, что вероятность "попаданий" в группе экспертов превышала вероятность срабатывания в режиме "ложных тревог". В целом, показатели чувствительности испытуемых данной группы находились в диапазоне от 0 до 1.7, причем большинство – в пределах значений 0.5–1.3 (см. кривые изочувствительности на

рис. 4). Отметим, что для эксперта ТО (большой музыкальный опыт и хороший интонационный слух) этот показатель оказался вторым по группе (d' = 1.2).

На рис. 4 показана и граница между положительными и отрицательными значениями установочного критерия c. Видно, что все полученные значения c положительны и находятся в диапазоне от 0.12 до 0.72. Следовательно, можно заключить, что тенденция к "пропуску" интонационных ошибок экспертами превалирует над тенденцией к объявлению "ложных тревог". Таким образом, эти данные, а также размещение индивидуальных показателей d' (нижний левый угол) дают основание предположить, что 20-центовый критерий, выбранный в качестве порога, несколько завышен.

ОБСУЖДЕНИЕ

Исследование было направлено на изучение особенностей восприятия и воспроизведения мелодических интервалов. Для его проведения были выбраны условия, обеспечивающие минимальный музыкальный контекст и естественную обстановку как для исполнения, так и для экспериментальной оценки чистоты вокальной интонации разными группами слушателей. Кроме того, были расширены рамки подхода к обработке полученных результатов.

Полученные экспериментальные данные позволили выявить тенденции, сходные с обнаруженными в музыкальной практике и опытах с изолированными интервалами. В первую очередь это тенденция к сокращению малых и расширению больших интервалов (Burns, 1999). Так, результаты выполнения вокального упражнения певцами показали, что малые секунды были исполнены в среднем на 11 и 6 центов (восходящий и нисходящий ряд) у́же, чем соответствующие им равномерно-темперированные эквиваленты (РТЭ). В то же время интервалы тритон и квинта, наоборот, превышали РТЭ на 10 и 3 цента (восходящие и нисходящие тритоны) и на 14 центов (чистые квинты обоих рядов).

Было показано также, что певцы воспроизводят малые (секунды) и консонантные (квинты) интервалы более точно, чем тритоны. Так, для секунд стандартное отклонение составило 10 (нисходящий) и 11 центов (восходящий ряд), для тритонов – 24 и 29 центов соответственно. Для чистых квинт эти значения занимали промежуточное положение – 16 и 22 цента. Полученные данные хорошо интерпретируются в терминах "прочности" интервалов и частоты их встречаемости в музыкальной практике, предложенных в работе Раковского (Rakowski, 1990). Так, известно, что малая секунда является одним из самых распространенных интервалов в западном музыкальном репертуаре, а тритон – одним из наиболее редких (Vos, Troost, 1989). Если предположить, что фактор встречаемости интервала определяет степень его "прочности", то наши результаты хорошо подтверждают этот порядок рангов. Так, разброс для малой секунды был значимо меньше, чем для остальных интервалов, а для квинты он был меньше, чем для тритона, который выступал в качестве наименее "прочного" из всех категорий использованных в эксперименте интервалов.

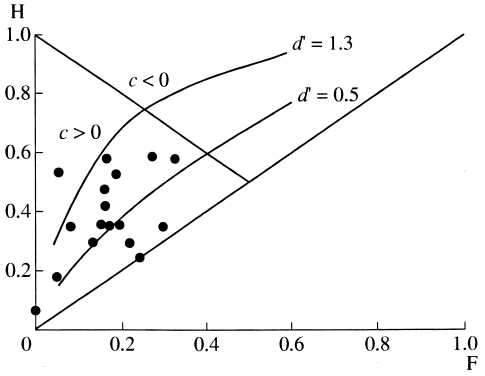

Рис. 4. Показатели чувствительности (d') и установки (c), полученные при анализе интонационных ошибок, выделенных группой экспертов.
Вертикальная ось (H) – коэффициент попаданий; горизонтальная ось (F) – коэффициент ложных тревог; кривые – изочувствительные линии (d' = const) (остальные пояснения в тексте).

Анализ полученных результатов с применением двух процедур сравнения позволил также рассмотреть вопрос о возможной стратегии настройки певца при воспроизведении интервалов. Сравнивались два варианта – настройка производится без обращения к контексту звуковысотного строя и при настройке учитывается контекст всей шкалы равномерной темперации с фиксированным стандартом высоты (А4). Отметим, что в реальной музыкальной практике возможны проявления обоих стратегий. Так, певцы могут воспроизводить мелодию как последовательность интервалов (интервал за интервалом) или ориентироваться на целостный образ шкалы с определенным положением по оси частот. Возможны ситуации и совместного их использования. Об этом, в частности, свидетельствуют личный опыт первого автора настоящей работы – профессионального певца, а также данные испытуемой ІО, обладающей абсолютным музыкальным слухом. Так, средние значения стандартного отклонения интервалов у этой певицы (исп. ІО) составили – 34 и 24 цента (стратегии настройки 1 и 2 соответственно). У остальных певцов данное соотношение было обратным – 22 и 34 цента. Можно предположить, что в отличие от других испытуемых певи-

ца IO использовала для коррекции интонации и абсолютный эталон высоты (440 Гц).

Однако в целом, полученные в работе результаты свидетельствовали о том, что певцы больше ориентировались на первый вариант настройки. Возможно, этому способствовал обедненный или квази-музыкальный характер исполняемых упражнений. Отметим, однако, что аналогичные условия встречаются и в реальной музыкальной практике, например, при чтении музыки с листа (Burns, Campbell, 1994) или тренировке слуха.

С точки зрения перцептивной оценки, анализ результатов восприятия вокальных интервалов показал значимое различие для двух условий прослушивания: текущая и отсроченная оценка интонации. Было показано, что испытуемые-певцы не способны адекватно оценивать чистоту своей интонации, выполняя вокальные упражнения. Полученные в этих условиях ответы близки к случайным, а различие между основными категориями оценок ("правильно" – "неправильно") не является статистически значимым. При прослушивании упражнений в записи процесс обнаружения интонационных ошибок (интервалов, исполненных "не в тон") значительно улучшается, и оценки становятся сопоставимы с результатами, полученными для группы экспертов. Так, средние отклонения от равномерной темперации, выделяемые группой певцов, равнялись 11 (оценки "в тон") и 25 центам (оценки "не в тон"). По группе экспертов они составили 8 и 31 цент соответственно. Таким образом, текущий сенсорный контроль за чистотой интонирования в процессе вокального исполнения можно было рассматривать как недостаточно эффективный. Точность перцептивной оценки интервалов певцами восстанавливалась только в условиях прослушивания записи своего вокального упражнения.

Точность оценки интонации зависит также от музыкального опыта и слуха человека (Гарбузов, 1948). Считается, что в максимальной степени интонационный слух развит у музыкантов, которые самостоятельно настраивают инструменты с нефиксированной высотой звука. Результаты работы подтверждают это положение. Так, показатель чувствительности d', определенный с помощью теории обнаружения сигналов, был одним из самых высоких у испытуемого – профессионального композитора (эксперт ТО), который обладал наибольшим музыкальным опытом. Кроме того, по мнению Гарбузова (1948), музыканты способны различать до трех градаций (зон) внутри категорий интервалов. При недостатке опыта и музыкального слуха границы между зонами могут сливаться. В нашем случае это касалось испытуемых с показателем d', близким к 0. Они были не способны выявить различия интервалов в пределах ±40 центов.

Однако большинство результатов свидетельствовало о выделении зонных градаций. Так, "чистые" интервалы (100% оценок "в тон") никогда не отклонялись от РТЭ более чем на 20 центов. Проявилась и зависимость ширины "чистой" зоны от категории интервала. Были получены значимые различия между интервалами (в первую очередь относительно тритона), которые проявились при анализе как усредненных (по группе испытуемых), так и индивидуальных данных. Например, у испытуемого ТО (высокий показатель d') ошибки в оценке тритона составляли от 38 до 41 цента, а для других интервалов – не превышали 20–25 центов.

Анализ результатов восприятия вокальных интервалов, проведенный с помощью теории обнаружения сигналов, позволил выделить и ряд общих характеристик интонационной чувствительности слушателей, а также оценить выбор порогового критерия. Было показано, что у всех 17 экспертов показатели чувствительности находятся в пределах диапазона от 0 до 1.7, показатели установки положительны, тенденция к проявлению "ложных тревог" не выражена. Эти данные позволили считать выбор порогового критерия адекватным и сделать вывод о том, что интервалы (малая секунда, тритон, чистая квинта) при отклонении от равномерно-темперированного эталона до ±20–25 центов будут восприниматься слушателями как "чистые", т.е. исполненные правильно.

Системное описание данных, включающее оценку перцептивного критерия чистоты интонации, а также вывод о нарушении сенсорного контроля за интонацией во время вокального исполнения, выступили как наиболее значимые и приоритетные результаты проведенного исследования, расширяющие современные представления о закономерностях восприятия и воспроизведения музыкальных интервалов человеком.

Работа поддержана грантами Научного Фонда Эстонии (№ 4712) и РГНФ (№ 030600372).

СПИСОК ЛИТЕРАТУРЫ

Гарбузов Н.А. Зонная природа звуковысотного слуха. М.: Изд-во АН СССР, 1948. 84 с.

ASA. Acoustical terminology SI. N.Y.: Amer. Standards Assoc., 1960. V. 1. 50 p.

Burns E.M. Intervals, scales, and tuning // The psychology of music / Ed. D. Deutsch. San Diego: Acad. Press, 1999. P. 215–264.

Burns E.M., Ward W.D. Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals // J. Acoust. Soc. Am. 1978. V. 63. P. 456–468.

Burns E.M., Campbell S.L. Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception // J. Acoust. Soc. Am. 1994. V. 96. P. 2704–2719.

Frances R. The perception of music. Hillsdale, NJ: Erlbaum, 1987. 375 p.

Goldstein J. An optimum processor theory for the central formation of the pitch of complex tones // J. Acoust. Soc. Am. 1973. V. 54. P. 1496–1516.

Green P.C. Violin intonation//J. Acoust. Soc. Am. 1937. V. 9. P. 43–44.

Hagerman B., Sundberg J. Fundamental frequency adjustment in barbershop singing // J. Res. Singing. 1980. V. 4. P. 3–17.

Hall D.E., Hess J.T. Perception of musical interval tuning//Mus. Percept. 1984. V. 2. P. 166–195.

Kagen S. On studying singing. N.Y.: Rinehart, 1950. 119 p.

Lindgren H., Sundberg J. Grundfrekvensforlopp och falsksang. Stockholm: Stockholm Univ., Inst. Musicology, 1972. 138 p.

Loosen F. Intonation of solo violin performance with reference to equally tempered, Pythagorean, and just intonations // J. Acoust. Soc. Am. 1993. V. 93. P. 525–539.

Macmillan N.A., Creelman C.D. Signal detection theory: a user guide. Cambridge: Cambridge Univ. Press, 1991. 492 p.

Moore B.C.J. Hearing. London: Acad. Press, 1995. 135 p.

Moore B.C.J. An introduction to the psychology of hearing. London: Acad. Press, 2000. 373 p.

Moran H., Pratt C.C. Variability of judgments of musical intervals // J. Exp. Psych. 1926. V. 9. P. 492–500.

Morrison S.J., Fyk J. Intonation // The science and psychology of musical performance / Eds R. Parncutt, G. McPherson. Oxford: Oxford Univ. Press, 2002. P. 183–198.

Mürbe D., Pabst F., Hofmann G., Sundberg J. Effects of a professional solo singer education on auditory and kinesthetic feedback – a longitudinal study of singers' pitch control // J. Voice. 2004. V. 18. P. 236–241.

Pierce J.R. The nature of musical sound // The psychology of music / Ed. D. Deutsch. San Diego: Acad. Press, 1999. P. 1–23.

Rakowski A. Intonation variants of musical intervals in isolation and in musical contexts // Psych. Mus. 1990. V. 18. P. 60–72.

Rakowski A., Miskiewicz A. Deviations from equal temperament in tuning isolated musical intervals // Arch. Acoust. 1985. V. 10. P. 95–104.

Rosner B.S. Stretching and compression in the perception of musical intervals // Mus. Percept. 1999. V. 17. P. 101–114.

Ross J. Measurement of melodic intervals in performed music: some results // Symposium: computational models of hearing and vision (summaries) / Ed. J. Ross. Tallinn: Estonian SSR Acad. Sci., 1984. P. 50–52.

Rossing T.D. The science of sound . MA: Addison-Wesley Publ. Comp, 1990. 686 p.

Shonle J.I., Horan K.E. The pitch of vibrato tones // J. Acoust. Soc. Am. 1980. V. 67. P. 246–252.

Srulovicz P., Goldstein J.L. A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum // J. Acoust. Soc. Am. 1983. V. 73. P. 1266–1276.

Sundberg J. Effects of the vibrato and the singing formant on pitch // Musicol. Slovaca. 1978. V. 6. P. 51–69.

Sundberg J. The science of the singing voice. IL: Northern Illinois Univ. Press, 1987. 216 p.

Sundberg J. The science of musical sounds. London: Acad. Press, 1991. 237 p.

Sundberg J., Prame E., Iwarsson J. Replicability and accuracy of pitch patterns in professional singers // Vocal fold physiology: controlling complexity and chaos / Eds P.J. Davis, N.H. Fletcher. San Diego: Singular Publ. Group, 1996. P. 291–306.

Terhardt E. Oktavspreizung und Tonhohenverschiebung bei Sinustonen // Acustica. 1969. V. 22. P. 348–351.

Terhardt E. Pitch, consonance and harmony // J. Acoust. Soc. Am. 1974. V. 55. P. 1061–1069.

Terhardt E. Intonation of tone scales //Arch. Acoust. 1988. V. 13. P. 147–156.

Terhardt E. Akustische Kommunikation – Grundlagen mit Horbeispielen. Berlin – Heidelberg: Springer, 1998. 505 p.

Ternström S., Sundberg J., Collden J. Articulatory perturbation of pitch in singers deprived of auditory feedback // Proc. Stockholm Music Acoust. Conf. SMAC-83. / Eds Askenfeldt, S. Felicetti, E. Jansson, J. Sundberg. Stockholm: Royal Swedish Academy of Music, 1983. V. 1. P. 291–304.

Titze I.R. Principles of voice production. Needham Heights, MA: Allyn and Bacon, 1994. 354 p.

Vos P.G., Troost J.M. Ascending and descending melodic intervals: statistical findings and their perceptual relevance // Mus. Percept. 1989. V. 6. P. 383–396.

Ward W.D. Subjective musical pitch // J. Acoust. Soc. Am. 1954. V. 26. P. 369–380.

Ward W.D. Musical perception // Foundations of modern auditory theory / Ed. J. Tobias. N.Y.: Acad. Press, 1970. P. 405–447.

Perception of Vocal Musical Intervals

A. Vurma*, J. Ross[1], E. A. Ogorodnikova[2]

Estonian Academy of Music and Theatre, Rävala puiestee 16, Tallinn 10143
[1]Also at University of Tartu, Ülikooli 18, Tartu 50090
[2]Pavlov Institute of Physiology, RAN, nab. Makarova 6, St.-Petersburg 199034

This paper reports two experiments. In the first experiment, 13 professional singers performed a vocal exercise consisting of three ascending and descending melodic intervals: minor second, tritone, and perfect fifth. Seconds were sung more narrowly but fifths more widely in both directions, as compared to their equally tempered counterparts. Standard deviation was the largest for tritones in performance, while seconds and fifths were intoned more similarly in size. In the second experiment, intonation accuracy in performances recorded from the first experiment was evaluated in a listening test. The performers themselves evaluated their performance almost randomly in the immediate post-performance situation but acted comparably to the independent group after having listened to their own recording. Data suggest that, due to the categorical nature of perception, melodic intervals may on average be 20 to 25 cents out of tune and still be estimated as correctly tuned by expert listeners.

Key words: perception, pith, musical intervals, musical temperation.

IV: Vurma, A. & Ross, J. (2007).

Timbre-induced pitch deviations of musical sounds.

Journal of Interdisciplinary Music Studies, 1: 33–50.

journal of interdisciplinary music studies
spring 2007, volume 1, issue 1, (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 edildmiş 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 free-lance 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.

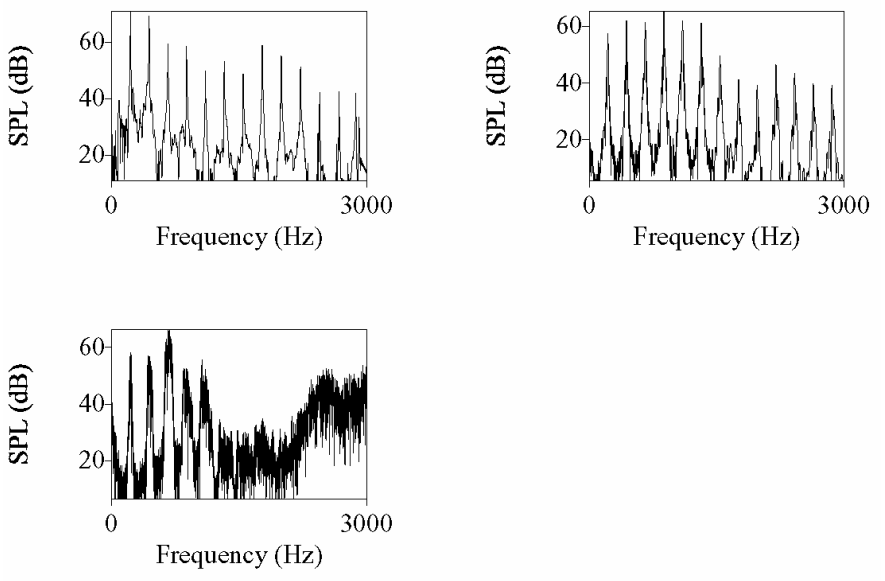

Figure 1. FFT spectra of sounds used in experiments. Top left: piano; top right: oboe; bottom: singing voice.

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.

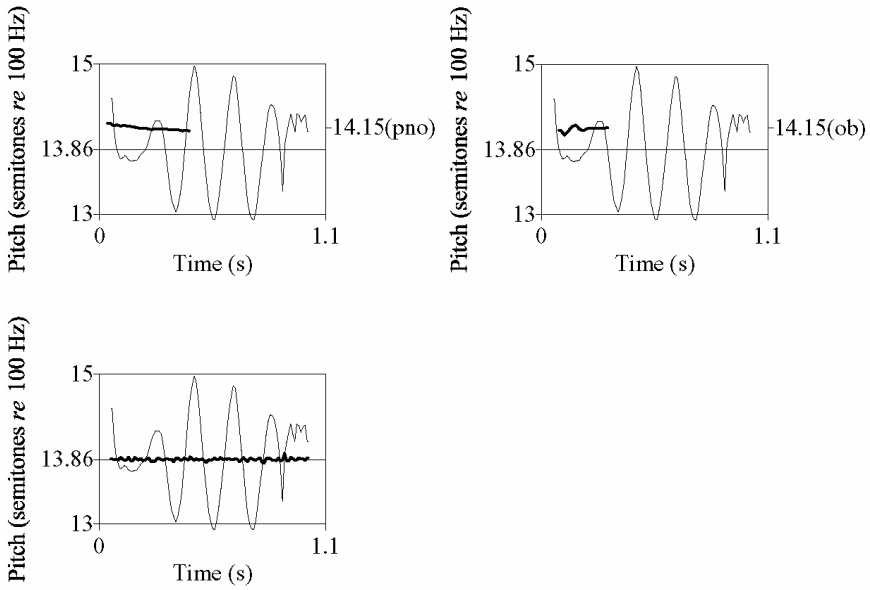

Figure 2. The fundamental frequency contours of sound pairs. Top left: piano (bold) and singing voice, sound A; top right: oboe (bold) and singing voice, sound A. Both panels represent the best match between the vocal and instrumental sounds in the second experiment, so that F0 of the former is 13.86 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).

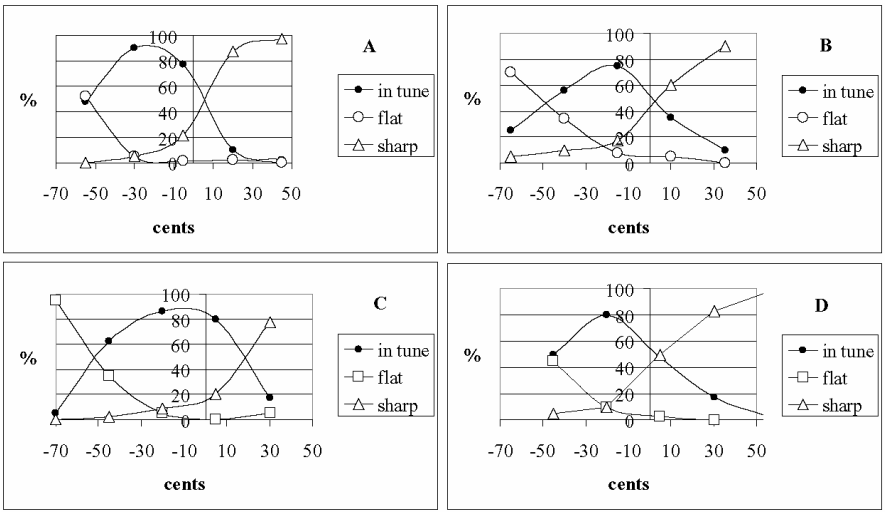

Figure 3. Distribution curves of 'in tune', 'flat' and 'sharp' ratings after comparison of the vocal sounds to the synthesized piano sounds. Horizontal axis: fundamental frequency of the vocal sound minus fundamental frequency of the instrumental sound; vertical axis: percentage of 'in tune', 'flat' 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 at bottom paragraph of the previous page) remain between plus 20 to plus 35 cents, and minus 53 to minus 69 cents, covering a region of 73 to 99 cents.

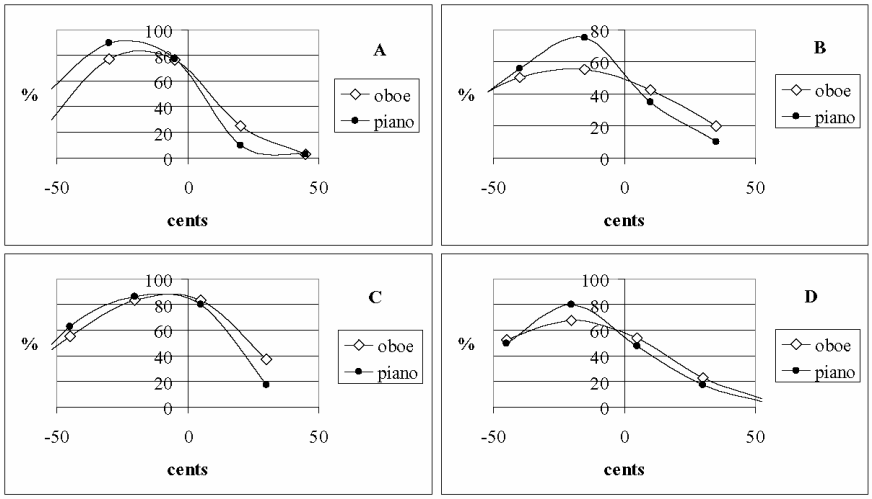

Figure 4. Distribution curves of 'in tune' ratings after comparison of the vocal sounds to the synthesized piano and oboe sounds. Horizontal axis: fundamental frequency of the vocal sound minus fundamental frequency of the instrumental sound; vertical axis: percentage of 'in tune' ratings. Top left: 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. approxi-

mately 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.

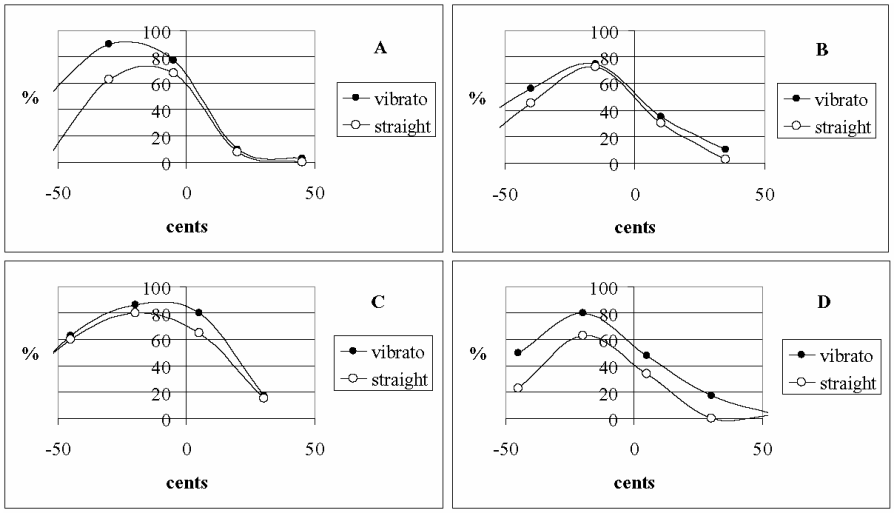

Figure 5. Distribution curves of 'in tune' ratings after comparison of the vocal sounds with and without vibrato to the synthesized piano sounds. Horizontal axis: fundamental frequency of the vocal sound minus fundamental frequency of the instrumental sound; vertical axis: percentage of 'in tune' ratings. 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 indi-

vidual 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 8th 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.