VETENSKAP OCH KONST

KUNGL TEKNISKA HÖGSKOLAN

Computer methods for voice analysis

Svante Granqvist

TRITA–TMH 2003:2
ISSN 1104-5787
ISBN 91-7283-461-7

Doctoral thesis
March 28, 2003
KTH Stockholm, Sweden

Till Åsa

KUNGL
TEKNISKA
HÖGSKOLAN

Computer methods for voice analysis

Svante Granqvist

Doctoral dissertation, Stockholm, March 28, 2003
KTH
Department of Speech, Music and Hearing

Abstract

This thesis consists of five articles and a summary. The thesis deals with methods for measuring properties of the voice. The methods are all computer-based, but utilise different approaches for measuring different aspects of the voice.

Paper I introduces the Visual Sort and Rate (VSR) method for perceptual rating of voice quality. The method is based on the Visual Analogue Scale (VAS), but simultaneously shows all stimuli as icons along the VAS on the computer screen. As the listener places similar-sounding stimuli close to each other during the rating process, comparing stimuli becomes easier.

Paper II introduces the correlogram. Fundamental frequency F_0 sometimes cannot be strictly defined, particularly for perturbed voice signals. The method displays multiple consecutive correlation functions in a grey scale image. Thus, the correlogram avoids selecting a single F_0 value. Rather it presents an unbiased image of periodicity, allowing the investigator to select among several candidates, if appropriate.

Paper III introduces a method for detection of phonation to be utilised in voice accumulators. The method uses two microphones attached near the subject's ears. Phase and amplitude relations of the microphone signals are used to form a phonation detector. The output of the method can be used to measure phonation time, speaking time and fundamental frequency of the subject, as well as sound pressure level of both the subject's voicing and the ambient sounds.

Paper IV introduces a method for Fourier analysis of high-speed laryngoscopic imaging. The data from the consecutive images are re-arranged to form time-series that reflect the time-variation of light intensity in each pixel. Each of these time series is then analysed by means of Fourier transformation, such that a spectrum for each pixel is obtained. Several ways of displaying these spectra are demonstrated.

Paper V examines a test set-up for simultaneous recording of airflow, intra-oral pressure, electro-glottography, audio and high-speed imaging. Data are analysed with particular focus on synchronisation between glottal area and inverse filtered airflow. Several methodological aspects are also examined, such as the difficulties in synchronising high-speed imaging data with the other signals.

Key words: voice analysis, perceptual analysis, fundamental frequency, correlogram, aperiodicity, Fourier analysis, high-speed imaging, laryngoscopy, vocal fold vibration, voice accumulation.

Table of contents

List of papers

Paper I

The Visual Sort and Rate method for perceptual evaluation in listening tests
Svante Granqvist
Submitted for publication

Paper II

The Correlogram: a visual display of periodicity
Svante Granqvist and Britta Hammarberg
In review

Paper III

The self-to-other ratio applied as a phonation detector for voice accumulation
Svante Granqvist
TMH-QPSR, KTH, Vol 45 2003, Submitted for publication

Paper IV

A method of applying Fourier analysis to high-speed laryngoscopy
Svante Granqvist, Per-Åke Lindestad
Journal of Acoustic Society of America (JASA) 110 (6) 3193-3197 Dec 2001

Paper V

Simultaneous analysis of vocal fold vibration and transglottal airflow; Exploring a new experimental set-up
Svante Granqvist, Stellan Hertegård, Hans Larsson, Johan Sundberg
Accepted for publication in Journal of Voice

Author's contributions to the studies

Paper I:	This work was carried out entirely by the author SG.
Paper II:	Author SG has invented this method, performed the computer programming, and written the major part of the manuscript. Co-author BH supplied most of the natural voices and is responsible for the choice of perceptual voice terms and also participated in revising the manuscript.
Paper III:	This work was carried out entirely by the author SG.
Paper IV:	The major part of this work was performed by the author SG. Co-author PÅL provided the high-speed recordings and participated in editing the manuscript.
Paper V:	Analysis of data and development of the analysis method was performed by co-author SG. Area measurement was performed by HL. JS supplied the idea for the experiments, SH performed the recordings in co-operation with SG, HL and JS. The manuscript was prepared mainly by SG and JS.

Preface

The path to the completion of this thesis has not been straight. I started in 1989 at the department of Speech, Music and Hearing (TMH), at the time called the department of Speech Communication and Music Acoustics. This was directly after receiving my degree in electrical engineering at the Royal Institute of Technology (KTH). My first project was to develop an interface card for PC computers that would transfer digital audio from DAT recorders into the computer. In those days, the main reason for this was the higher quality of the analog-to-digital converters in the DAT than in the computers. When that project was finished, I was asked if I could “install Windows on a few computers and help people out a little”, something that after a while became more formalised as the responsibility for the PC computers at TMH. I continued this work, at least part time, until the beginning of 1997, when I decided to pay full attention to my doctoral studies. In 1991, I also started teaching the electroacoustics course, first as a laboratory assistant, later running the tutorials, and last year (2002) assuming the responsibility for lectures. It might be thought that these activities have hampered my development as a researcher, but I actually think it is the other way around. My time as a system administrator taught me what users expect from computer programs, and also what could go wrong, both indispensable insights for a programmer. Also, my time as a teacher forced me to understand the topics I teach from more than one viewpoint, and it has also made me more comfortable with speaking in front of people.

My research career started somewhere in the mid ‘90s when I was involved in studies of voice carried out by Jan Gauffin, Britta Hammarberg and Stellan Hertegård. My first stumbling attempts in this field involved perturbation extraction and synthesis, and pretty soon I developed an interest for additive synthesis as a mean to synthesise subharmonics. Also, thanks to Britta Hammarberg, I started examining the relations between perception and acoustics. My experiences of listening to synthesised voice qualities made me realise the great difficulties for subjects asked to rate voice stimuli. I realised that part of the scatter in perceptual data was due to the test situation. This eventually lead to the work that is presented in paper I.

Most of the work presented in the other articles has a similar origin, in the sense that I rarely could predict the outcome when I started. This approach caused me to visit several dead ends, which may appear as a waste of time. On the other hand, if only the projects with an entirely predictable outcome would have been initiated, I would probably have missed some innovation.

I have been involved in some publications not included in this thesis, see Appendix I. They were excluded either because I was not the driving force in the projects, or because they are not directly related to voice. Appendix II lists some of the computer programs that I wrote. Some, e.g., LarVib, have been developed directly for investigations included in the thesis. Other, such as DeCap, do not represent novel research, but have still been extensively used in these investigations.

To make progress, both personally as a scientist and for research in general, successful projects have to be published. In the beginning of my doctoral studies this was not my primary concern and as most newcomers to the research, I needed guidance in the art of writing scientific articles. After fully realising the importance of publishing I had great help of Johan Sundberg to put my work on paper. He has been a great tutor, especially, but not only, in this regard. So, if I were to give advice to a new doctoral student today, it would be to get published and to have good and detailed guidance in writing the papers.

When I look back on my days at TMH I see a lot of expertise, friendliness and open minds. The department has offered a great research environment for my doctoral studies, and I hope to keep working there for a long time still.

Svante Granqvist, February 2003

Introduction

Description of voice qualities is an important field to many people. Voice properties are important to the everyday social interaction between people. In voice quality research, the two extremes that are encountered in singing and in pathological voice have attracted much attention. In terms of technology for quantifying voice quality, these two extremes have many things in common.

The voice has primarily been studied in three domains. One is by perceptual analysis. This is frequently utilised on an informal basis in most people's everyday life, but also in the voice clinics as a first assessment of a patient's vocal status. The perceptual rating of voice quality can also be formalised. Such formalisation can consist of development of a set of sound-describing adjectives that are found relevant for voice quality. In such cases, the listeners have to be trained and "calibrated" in a continuously ongoing process. Also, they need efficient tools for performing their perceptual ratings.

Voice quality can also be studied with respect to its acoustic and aerodynamic properties. The signal that leaves the mouth is recorded by means of a microphone or a Rothenberg mask. This type of analysis has potential for detecting voice characteristics that relate perception to physiology. It also has the advantage of being less invasive and costly than most physiological examinations.

Finally, the voice can be examined in the physiological domain, where the vocal folds and the vocal tract are studied with laryngoscopy or other types of imaging techniques such as x-ray, magnetic resonance or ultrasound, and also electroglottography, photoglottography etc.

All these three domains of studying voice have contributed importantly to the present knowledge regarding voice production. Also, each of these domains has developed a terminology of its own. These different terminologies are of course related, but by no means identical (Granqvist, 2000). For example, the perceptual correlate of fundamental frequency cannot be equated with fundamental frequency and should hence be denoted pitch. Likewise, the degree of phonatory pressedness/hyperfunction is related to but not identical with the closed quotient of the vocal fold oscillation.

The present thesis is mainly focused on computer methods for analysis of voice. The different methods can be regarded as tools for voice measurements in all these three domains.

Paper I

Perception of voice is an important aspect of voice quality. Perception can be related to physiological and acoustic properties, but there is rarely a simple relation between these domains. Therefore perceptual rating plays an important role in voice research.

There is, however, normally a rather large spread in the data obtained from perceptual tests. Different methods for reducing this spread have been examined in the literature. One example is training of the listeners in terms of continuous feedback from colleagues and also through courses in the education. Such training has lead to several more or less standardised protocols for voice rating, for example GRBAS (Ishiki et al, 1969) or SVEA (Hammarberg, 2000) developed for slightly different purposes. Selecting the number of parameters is problematic; a too large number of parameters may cause different subjects to categorise the voices under different parameters, yielding very different outcome from the subjects even if they do not really disagree on the voice quality. On the other hand, using too few parameters might force the listeners to rate entirely different qualities along the same scale. Therefore, it appears advantageous to select a number of parameters relevant to the material in pilot experiments before the tests are performed, since this should help in optimising the number of parameters for the stimuli chosen for the test. Also the choice of rating scale, e.g., equal appearing interval scales (EAIS) or visual analogue scales (VAS) is important (Wewers & Lowe, 1990).

Apart from training of subjects and selecting relevant parameters there are also methodological aspects regarding internal and external references that will affect the outcome of a perceptual test (Gerrat et al, 1993; Berliner et al, 1978). During a voice-rating test, the listener might be assumed to use his/her previous knowledge of voice quality, but he/she could also be assumed to rate the voices in relation to other stimuli presented during the test. In other words, the subject can be assumed to use both internal and external references. The only way of assuring that rating of stimuli is performed in relation to internal references is to rate only a single stimulus in each test. To some extent, this can be said to take place in the voice clinic when the logoped or physician listens to a single patient and perhaps makes notes about this in the patient's file. The opposite case would appear if stimuli were to be compared to fixed anchor stimuli supplied as a reference. In this case, internal references can be assumed to play only a minor role in the outcome of the result. However, relevant anchor stimuli might be hard to obtain, since the perception of voice parameters are interleaved. For example, the perception of vocal fry might be influenced by the F_0 of the particular sound sample. It is generally thought that external references are more stable than internal references, but the problem of obtaining valid anchor stimuli has so far hampered the development of standardised voice rating using anchor stimuli. Another way of utilising external references is to organise stimuli in all possible pairs, and to compare the voice qualities within those pairs. This approach is problematic due to the large number of pairs resulting even for a relatively few stimuli.

Paper I introduces the Visual Sort and Rate (VSR) method for voice rating. Each stimulus is represented by an icon on the computer screen and the listener drags these icons and places them along a VAS according to their perception. During the rating process, similar-sounding stimuli will become positioned near each other, and thus comparison between those becomes simplified. In a sense, then, each stimulus can be said to serve as anchor stimuli for the other, and thus the results obtained are relative to the remaining stimuli in the test, rather than to the internal references of the listeners.

Paper I also compares this method to VAS on paper and VAS implemented as a scrollbar on a computer. Correlation coefficients between sessions are statistically examined. In a single-parameter test, correlation coefficients were significantly higher for the VSR method, than for both of the other methods. Typically, the correlation coefficients increased from about 0.93 to about 0.98. The advantage of using the VSR method for tests involving more than one parameter seems smaller. In this case, only one significant difference was observed, and this difference was to the advantage of the VSR method as compared to the VAS on paper. This aspect needs further analysis in future research.

Paper II

Fundamental frequency is another important parameter in voice analysis. It is closely related to the pitch. Obviously, pitch is important in the case of the singing voice, but also in the case of pathological voice, where F_0 and F_0 perturbation is assumed to play an important role. Many algorithms have been developed to measure F_0, which may be based on event detection (peak picking, zero-crossing), spectral or cepstral methods, or waveform matching (e.g. Hess, 1983). All these methods present a likely numerical value for the F_0 in the signal analysed. However, both the singing voice and the pathological voice present problems for all of these algorithms, since there are cases where the F_0 is not well defined. In the case of pathological voice, a multi-cyclic pattern of period times can often be seen. Strictly speaking, the fundamental of such a voice is an integer factor lower than the unperturbed F_0. However, if the perturbation is small, extracting the lower F_0 makes little sense. For the singing voice, a similar problem exists. For example, when the second partial coincides with the first formant, the waveform may appear to have a fundamental equal to that formant frequency, rather than to that of the fundamental. Thus, it is clear that extracting a numerical value for F_0 is sometimes a difficult or, for some pathological voices even an inappropriate task. In some cases, such as in the singing voice example mentioned above, it should be obvious for the experimenter that the true F_0 differs from the automatically extracted F_0. In those cases, it would be advantageous if the F_0 extraction program could present several alternatives for the F_0.

This is the strategy chosen in the correlogram method presented in paper II. The correlogram utilises the Pearson correlation between two moving time-windows of the signal. When the delay between the windows corresponds to one or more fundamental periods, a high correlation will be the result. This correlation is displayed in a grey scale, such that black corresponds to a correlation of 1 and white corresponds to correlations less than or equal to 0. In the graph, the start time of the first window is reflected by the x-axis while the y-axis corresponds to the delay between the two windows, mostly converted into frequency. The resulting graph is similar to a spectrogram, but displaying periodicity information instead of spectral properties.

For periodic signals, several candidates appear at F_0, $F_0/2$, $F_0/3$ etc. No automatic selection is made regarding what candidate represents F_0. The purpose of the graphs is twofold: First, the software can be equipped with tracing capabilities, in order to facilitate semi-manual F0 extraction. In this case, the user has the opportunity to select the appropriate candidate. Second, if the selection of a single candidate is inappropriate because of multiple periodicities, the graph in itself can be used as a description of the periodicities that appear in the signal.

In paper II, the correlogram is tested on synthetic signals as well as on some archetypical real voices, thus demonstrating its behaviour for some typical signal properties and correlogram settings. Several signal properties are visualised in a correlogram. For example, bi-cyclic signals are displayed with altering odd-order candidates, but with stable even-order candidates. Random period time variations are displayed with all candidates fluctuating, and no stable candidates. Voice signals with a well-excited first formant appear with pronounced side-bands while hypofunctionally breathy voices, which typically have a dominant fundamental, appear as wide candidates without side bands.

Paper III

Some types of voice pathology are related to vocal behaviour and occupational factors (Fritzell 1996). The prime method to deal with such voice problems is to make the patient alter this behaviour. In some cases the vocal habits may be assessed in the clinic, but in other cases the vocally abusing factors are not provoked in the usually quiet environment of the clinic. Environmental noise can be simulated (e.g. Neils & Yairi 1987; Ternström et al, 2002), and to some extent provoke patients to resort to their habitual detrimental voice techniques. Yet, there are cases when registration of the patient's vocal habits and acoustic environment during the normal conditions of work is required. Such recordings may be useful for single patients, but can also be used in research, for example to determine typical vocal loading and noise exposure for different occupations, and the related occupational voice health hazards.

Recording of habitual voice use in the normal environment entails some difficulties and several methods to acquire data have been tested in the literature. The vocal fold vibration can be detected either directly by means of a contact microphone (e.g. Ohlsson et al, 1989; Szabo et al, 2001; Masuda et al, 1993) or electro-glottography (EGG) (Kitzing, 1979). Also, it can be detected by recording the acoustic signal from the mouth (e.g. Buekers et al, 1995; Rantala et al, 1998; Popolo et al, 2002). These methods have both advantages and disadvantages; the direct methods depend on good contact and placement of the transducers on the patient’s neck. Beard, loose skin or subcutaneous fat have been shown to be problematic factors here (Szabo et al, 2001), as well as the patient’s clothing, particularly the collar. Also practical problems with the attachment of the microphone may occur. On the other hand, the direct methods provide a signal relatively free from environmental acoustic noise, and can thus be utilised fairly easily for detection of phonation.

Unlike direct recording methods, recording of the acoustic signals can be performed relatively independently of the anatomy of the patient’s neck. Also, with appropriate microphone placement, much of the problems with the patient’s clothing can be avoided. On the other hand, not only the voicing from the patient will be recorded, but also the environmental sounds. This can be both advantageous and disadvantageous; while environmental sounds may disturb the analysis, the acoustic characteristics of both the subject and the environment can be measured, provided that the patient’s phonation can be properly detected.

Thus, when using microphones, there is a need to detect instances of the patient’s phonation and also to make sure that the microphone signals are dominated by the patient’s phonation during those instances. Luckily, with microphones near the subject’s mouth, the signals will be dominated by the patient’s voice during phonation, since people tend to speak loudly in noisy surroundings.

Detection of phonation, however, is more problematic and has typically been performed by also using a contact microphone (which entails the practical problems previously mentioned) or to utilise two separate microphones. Those two microphones can be mounted in different ways, depending on the method for phonation detection. In the method by Airo and associates (Airo et al, 2000), one microphone is mounted near the mouth of the subject, and one is mounted reasonably far from the mouth. Phonation was assumed to have occured when the level at the mouth was sufficiently higher.

The method presented in paper III in this thesis uses two microphones attached near the patient’s ears, at equal distance from the mouth. Phonation detection is based on the amplitude and phase relationship between the two microphone signals. After the phonation detection, the original microphone signals are automatically sorted into separate channels reflecting the sounds from the subject and the environment. Post processing further improved the sorting. The following parameters could be estimated using the output from the method; phonation time, speaking time, F0 for the subject and SPL of both the subject and environmental sound.

The paper describes the technical part of an investigation of the vocal load of pre-school teachers at work. It describes the method as it was used during that study. Even though the method performed well, there is a potential for further improvement. The main detail that might be improved is the detection of soft phonation with few harmonics. The problem originates in that low frequencies appear approximately in-phase at the microphones, regardless of origin. In the present implementation, this problem was handled by giving the low-frequencies less weight than the higher frequencies. This was realised by adding a second order high-pass filter, which worked well in most cases. However, this method does not completely eliminate the fact that ambient low-frequency sounds will be falsely identified as phonation if there is a very low relative level at high frequencies. In future development of the method, a first order high-pass filter could instead be applied in the sum channel only. If the cut-off frequency is properly chosen, this would enable an approximately constant self-to-other ratio of 0 dB for all frequencies, in the presence of only ambient sound.

The method is presently implemented on a computer. This implies that the signals are first recorded on tape and the processing is performed off-line. This has been advantageous during the development of the method and has also allowed for individual adjustments of analysis parameters for the different subjects and also verification of results. However, the procedure of transferring the tapes to the computer and thereafter running the signals through the analysis program, is quite elaborate. For clinical purposes it would be better to have an implementation that performed the analysis on-line, while the recording is being made. Such a device could give quick access to averages of the extracted parameters for parts of or the entire working days, or weeks. The current development of consumer electronics may offer interesting opportunities, since hand-held computers are getting ever so small and light-weighted (Popolo et al, 2002). The main obstacle in the presently available devices is the number of analogue inputs, which is mostly limited to one and the capacity of the batteries. With the increasing memory capacities, and also with the possibilities for audio compression, it might even be possible to save the audio signal for short excerpts, thus allowing verification of proper phonation detection.

Paper IV

High-speed imaging of the vocal folds is a technique that is becoming more commonly available. Contrary to stroboscopy, it can be used to visualise non-periodic vibrations and also vibrations that occur at more than one frequency, and this property is important for the investigation of the vibratory patterns that occur in some voice pathologies. However, the large amount of data that is generated sometimes makes the analysis and recording time-consuming. For example, a four second recording at 2048 frames per second and a resolution of 256 by 64 pixels, 8 bits per pixel, takes more than five minutes to view if played back at 25 frames per second and occupies 128 MiB of hard disk space. Obviously, management and analysis of such amounts of data is overwhelming.

In spite of these disadvantages, high-speed imaging of the vocal folds has become an important research tool and is also sometimes used for special cases in the voice clinic. The analysis methods for the recordings include visual inspection of the image sequence, area measurement (e.g. Larsson et al, 1999), kymography (Švec, 2000; Neubauer & Mergell, 2000) and high-speed glottography (HGG). All of these methods except the first reduce the spatial information thus introducing the risk that vibrations in parts of the image be left out. The visual inspection, on the other hand, requires moving playback, if vibratory patterns are to be observed. Therefore, this method is not appropriate for publication in printed form. It might also be hard to determine temporal relations of the vibrations in the image, in particular if the vibrations involve several frequencies.

Paper IV presents a new method for analysis of high-speed recordings. The analysis is based on the fast Fourier transform (FFT) of the light intensity variation in each pixel of the image. For example, if the images have a resolution of 256x64 pixels, 16384 waveforms and their FFTs are extracted, one for each pixel in the images. The results can be displayed in several ways, among other by introducing colouring of the image. The frequency of interest, usually the fundamental of the oscillation, is selected from the Fourier transform manually, and the amplitude at that frequency is displayed in terms of colour on top of a single image from the sequence. In this way, information regarding the oscillation can be presented without reduction of the spatial information. If the phase information from the Fourier transformation is reflected in terms of colour hue, the phase relations of the oscillations can be illustrated as well, and not only in the left-right dimension as in the kymogram. It is also useful in detecting co-vibrations in the ventricular folds or other parts of the larynx. These vibrations can occur at the fundamental frequency of the glottis, or at other frequencies. Cases have been observed where vibrations in the ary-epiglottic folds gradually increase in amplitude and finally affect the vocal fold vibrations such that the vocal fold vibrations ultimately become chaotic.

Some care must be taken since the time-varying light intensities are sampled at a rather low sampling rate, and without a proper anti-aliasing filter. This means that aliased harmonics can sometimes be observed at frequencies that do not actually appear in the original signals. However, bearing this in mind, these effects can mostly be disregarded.

Another problem with the method involves the fact that the transfer function from mechanical vibration to light-intensity oscillation is non-linear. This implies that higher harmonics from the Fourier analysis are of relatively little relevance for the mechanical oscillation.

Also, if several frequencies of vibration occur in the same pixel, there will be intermodulation effects that are sometimes hard to predict. However, when such vibrations of different frequencies occur in different pixels in the image, the problem with intermodulation becomes smaller, and the method has shown to be a valuable tool for finding the origins of the oscillatory frequencies in, for example, the biphonic phonation shown in figure 1. This patient was diagnosed as having a Reinke's edema and the glottis was effectively split in two parts. In the audio signal, two separate frequencies can be seen at approximately 350 and 400 Hz. The spectral analysis reveals that the two parts of the glottis oscillate at those two separate frequencies. The features of this phonation would have been hard to illustrate using any of the other available techniques.

The method presented in paper IV was also utilised in Paper V where it was combined with kymography.

Paper V

Voice production can be analysed in many different aspects, and there are several physical phenomena that contribute to the emitted sound. The parameters include the airflow in the glottis, cross-sectional glottal area, contact area between the vocal folds, vocal tract shape, and subglottal pressure. These parameters are often studied separately, but there is also important information in the synchrony between them. Thus, simultaneous recording of several parameters may increase the understanding of voice production.

The synchrony between parameters, for example between glottal area and airflow, is one important factor. From models we know that the airflow should lag behind the glottal area, partially due to the inertia of the air within the glottis. Compared to the area waveform, the flow waveform will thus be skewed to the right. This effect is important since it will cause the flow derivative to increase in the closing phase, and it has been shown that the flow derivative is strongly correlated to radiated acoustic power (Fant, 1982).

In paper V, simultaneous recording of several parameters were performed. One aim of the study was to examine the test set-up and to explore the problems that may occur when a large number of parameters are recorded simultaneously. Another aim was to investigate the difficulties with synchronisation between the high-speed imaging and the other signals. The present test set-up involves two separate clocks for sampling; one in the high-speed camera and the other in the multi-channel tape recorder that was used for the analogue signals. These clocks were synchronised in the present study by a simultaneous recording of the foot pedal that arrested the high-speed recording. However, as there was some uncertainty regarding the exact clock rates, only the data close to the synchronisation pulse turned out to be useful. Furthermore, the timing between the synchronisation pulse and the last recorded frame was examined down to a sub-frame level.

Finally, the recorded oral airflow was manually inverse filtered using the DeCap computer program. This program also displayed the EGG signal as an aid to determine the instant of glottal closure. The resulting estimate of the glottal airflow was then related to the glottal area derived from the high-speed recordings in terms of Lissajou figures since such figures are particularly well suited to illustrate small phase differences.

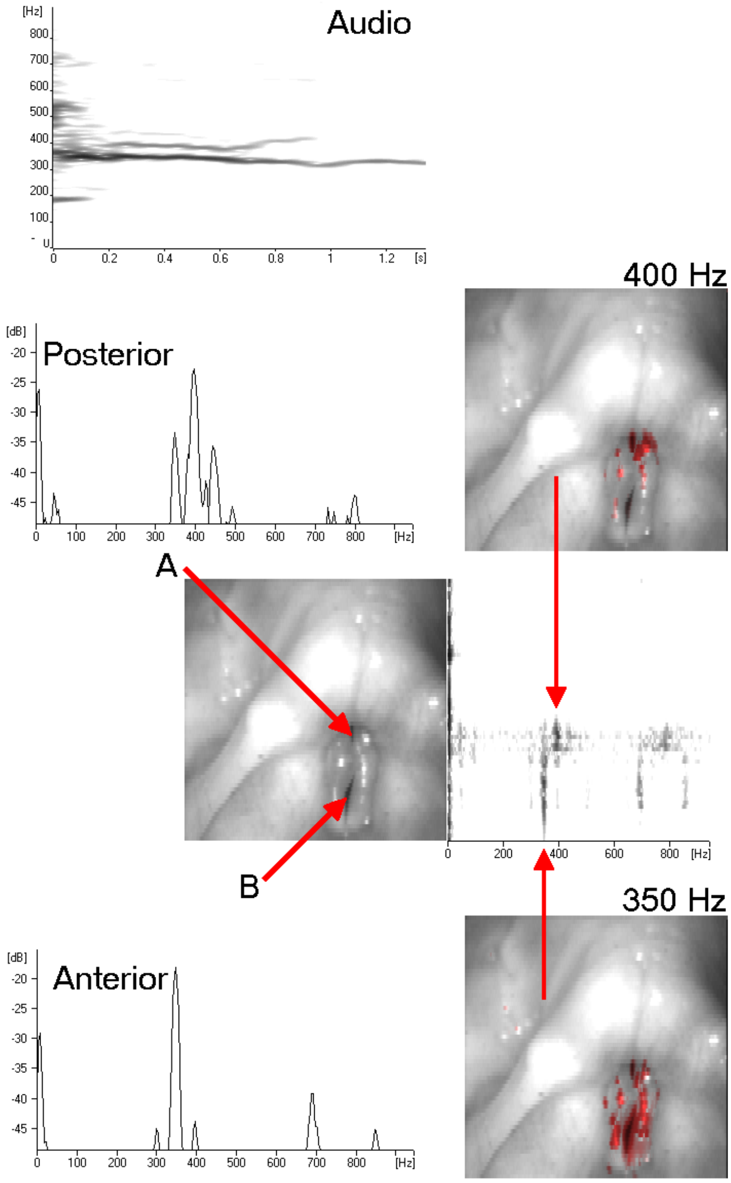

Figure 1. Example of biphonia. The top panel is a spectrogram of the audio signal indicating independent oscillations at two different frequencies.

Second row; spectrum of the light intensity oscillations, at the pixel indicated by arrow A, in the posterior part of the glottis; the right image shows the light intensity oscillations at 400 Hz over the entire image.

Third row; line-average spectrum of the light intensity oscillations in the different lines of the image shown to the left. Dominating oscillations can be seen at 400 Hz in the lines corresponding to the posterior part of the glottis, and dominating oscillations at 350 Hz occur in the lines corresponding to the anterior part of the glottis.

Last row; spectrum of the light intensity oscillations, at the pixel indicated by arrow B, in the anterior part of the glottis; the right image shows the light intensity oscillations at 350 Hz over the entire image.

See text for more information. (From "Fourier Analysis of High-Speed Laryngoscopic Image Sequences", oral presentation at Care of the Professional Voice, Philadelphia 2000 28/6-2/7).

Because of the limited accuracy of the synchrony only a small part of the recorded material could be analysed. Yet, the data confirmed that the glottal airflow lags behind the glottal cross-sectional area. It also gave some visual evidence of a piston movement of the vocal folds during the closed phase in conjunction with a non-zero glottal flow. The method appears promising but should be further developed, e.g., by solving the synchrony problems.

Future studies could include even more parameters than those of paper V. To quantify the piston movement of the vocal folds during the closed phase, absolute calibration of the glottal dimensions would be necessary and could be accomplished by including laser triangulation. To determine the dynamic transglottal pressure drop, subglottal pressure should be measured by means of tracheal puncture. In any event, recordings with high time-resolution and including multiple parameters crucial for the calibration of computer models of the vocal fold oscillator. With the help of such recordings, detailed information on the time-varying forces that act on the vocal folds during phonation could be acquired.

Discussion

The present thesis consists of a sequence of complementary investigations, each of which presents a method for computer based analysis of voice. The investigations do not form a coherent sequence. Rather they emerged from attempts to meet specific needs that occurred in the interaction with colleagues at TMH and other departments. The common denominator is that they concern computer methods for voice analysis and there are some direct links between the studies. For example, the Fourier method described in paper IV was used in paper V. However, to a large extent each investigation represents a stand-alone work and therefore has been related to current research separately above. This discussion will consider the topic of voice analysis in general terms and not always in direct connection to the presented papers.

Voice research is and must be multidisciplinary and voice is best understood if studied from several aspects. Thus the articles in this thesis represent a wide range of methods, including perception, acoustics and physiology. They are aimed at solving some of the problems that occur in characterisation of voice and have in common that they represent tools for analysis, rather than analyses per se. The problems addressed have all occurred in the authors' real-life. They do not necessarily form a coherent group of problems, but are rather the result of needs in the research environment of TMH and associated departments.

Perceived voice quality is surrounded by a more or less well-founded set of adjectives. For untrained listeners, the meaning of these adjectives is usually not very stable, at least if regarded as measurement instruments. To achieve stable perceptual rating of voice quality, the methods have to be standardised and calibrated, just like any other instrument that performs measurement. Admirable work in this field has been carried out in several contexts, in particular in voice clinics and in the education of voice professionals. Yet, differences remain. In part, this may depend on language differences. The requirement to share information in the scientific community in terms of written language would be another reason for differences. Here, the recent development of audio technology for computers and the Internet could play an important role for standardising and calibrating perceptual methods for voice quality measurement. For example, a single Internet site could be used to make archetypical voices available to the entire voice community. Following the example of Internet discussion forums, the community could supply complementary stimuli, aid rating of stimuli, and also discuss and even decide whether particular stimuli are appropriate as archetypes or not. This could eventually result in a commonly accepted and well-defined voice terminology.

Another way of calibrating perception of voice quality is to relate the descriptions to physical entities that can be objectively measured. This is sometimes problematic since there rarely is a one-to-one relationship between physics and perception. Nevertheless, tying voice adjectives to physics is important for achieving agreement on terminology and enhanced understanding of the mechanisms of voice production and pathology.

The technical methods also have to be subjected to critical review. The fact that a method is implemented on a computer does not necessarily mean that the results will be reliable or valid. Also, a method may yield valid results for some voices but completely erroneous values for other voices. Hence, it is important that the methods allow for examination of the accuracy of the results, not only during the development phase.

Unfortunately, in the field of pathological voice, computer methods sometimes are disturbed by a high amount of perturbation. In such cases, the researcher sometimes has to take a step back and go through the hardship of using manual or semi-manual methods rather than the completely automatic push-the-button methods.

The correlogram, binaural microphone technique and high-speed Fourier methods are all examples of methods leaving the final word to the researcher. The correlogram can be seen as

F_0 analysis without an automatic selection of F_0; the Fourier method analyses high-speed laryngoscopic imaging without detecting the edges of the glottis; the binaural microphone technique *does* make an automatic selection, but the results can easily be examined for accuracy and the last steps of extracting averages for F_0 and SPL are performed as a second step.

An interesting field that may receive increased attention in the near future is modelling of vocal fold vibration for pathological voices. This may allow the testing of surgical procedures before they are actually applied. Such simulations would be valuable for the individual patient in terms of more accurate predictions of the outcome of a surgery. They should also be useful in the education of surgeons and for development of new surgical procedures. Of course, the validity of the simulation is crucial. Multi-parameter measurements like in paper V should be useful in validating and calibrating such computer models.

Hopefully, some of the methods presented in this thesis will improve the understanding of voice production. However, for real progress the methods also have to be accepted among voice scientists. In order for this to happen, they must not only be presented in scientific articles, but also, computer programs implementing the methods must be made available. As many voice scientists do not have resources to perform computer programming, it is thus important for method developers to design well-behaved and user-friendly implementations of the methods. Even though the step to a final, user-friendly version often is large and difficult to finance, the recent development of programming tools and of the Internet have offered interesting opportunities with regard to this.

Acknowledgements

This work was supported by research grants from Swedish Research Council for Engineering, Bank of Sweden Tercentenary Foundation and Swedish Council for Working Life and Social Research (formerly Swedish Council for Work Life Research)

I also wish to thank everyone at the department of Speech, Music and Hearing at the Royal Institute of Technology and everyone working at the department of Logopedics and Phoniatrics at Huddinge University hospital for the great environment they are all part of. Having knowledgeable people around during the years I have spent with this thesis has been invaluable and this thesis would not have been the same without these people and the open-minded, prestige-free tradition that rules in both these departments.

Among these and others, there are some people that I particularly want to thank:

Jan Gauffin, my first supervisor, for all the great ideas and discussions during my early days as a doctoral student and for the days in his computer-crowded basement after his retirement.

Johan Sundberg, my co-author and second supervisor, who has lead me through the main struggle with publishing my papers and for giving me the deadlines that I so desperately need to finish my work. In my opinion, he represents a very good example of how a supervisor should aid his doctoral students through the ordeals of writing scientific papers.

Anders Askenfelt, who became my third supervisor at quite a late stage, but who has been around for discussions during my entire time as a doctoral student. I have also very much enjoyed our work with the electroacoustics course.

Johan Liljencrants for being the great guru of acoustics. … and everything. We miss you at the department, but know that you enjoy the freedom of your retirement.

Britta Hammarberg, my co-author, for her enthusiasm regarding voice qualities and for serving as a bridge between technical people like myself and the reality in the voice clinic. She should also be thanked for her never-ending additions to our articles, which has brought them to a standard that I never could have achieved by myself.

Per-Åke Lindestad, my co-author, for his well-trained vocal folds and for his skills with the endoscope. If vocal folds could smile, his would.

Stellan Hertegård, my co-author, for his laughs, his enthusiasm for new projects and for his gentle manoeuvring of the fiberscope during the sessions that I have had to serve as a subject myself.

Hans Larsson, my co-author, for all the tech talk at his office.

Maria Södersten and Annika Szabo for our work together in the past and the future and for the fun at conferences.

Joakim Westerlund for making the statistics seem understandable and for taking the time to analyse my data even though there was none.

Cathrin, the spiderwoman, for her editorial skills and for our chats in the car when going home from the TMH parties.

Markku, Caroline, Niclas, Peter and Payam, Ebi, Paolo. I hope you realise how important you are for me and everyone who work at TMH.

Lennart Neovius for bringing some of my programs to a bigger audience.

All the users of my computer programs. You might think that you are indebted to me, but it is actually the other way around. Without the feedback from you, this thesis would never have come about.

The subjects in my studies. It is amazing how much time some people have spent helping me without compensation.

The Formant Orchestra for proving that music is fun.

Sofia, for the many talks on work and life, and for so bravely writing articles together with me defying our incompatible writing styles. Monica, for her courage to work in the technical environment at KTH and for crashing my programs. Peta, for her organising skills and for offering to help arrange my dissertation party. Mikael for finally providing me with the acoustical resistance I need. I am looking forward to the coming time as roommates in the new building. Kjetil, for bringing my Norwegian to unexpected levels, Eva for her energy, Sten for all the new courses, Anders for his cool attitude, Roberto for half of his emails, Erwin for his chs, Erik for his knocks on wood, Eric for the long lunches, Alex for keeping watch over the department at nighttime and everybody in the music acoustics group for the great times I have had with you. You're the best!

My mother and my father and my big brothers Ulf and Åke for all the good times of my childhood and present.

Finally, and most of all, my children Håkan and Linus for having grown to be such nice persons and for starting to find your own paths, Anna och Oskar för att ni är så snälla och goa, and most of most of all, Åsa ♥ for the love and depth of our long-lasting relationship.

Appendix I

Selected publications not included in the thesis

Voice related

Szabo A, Hammarberg B, Granqvist S, Södersten M. Methods to study pre-school teachers' voice at work - Simultaneous recordings with a voice accumulator and a DAT recorder. Accepted for publication in *Log Phon Vocol.*

Södersten M, Granqvist S, Hammarberg B, Szabo A (2002). Vocal behaviour and vocal loading factors for preschool teachers at work studied with binaural DAT recordings. *Journal of Voice* 16/3: 356-371.

Lindestad P-Å, Södersten M, Merker B, Granqvist S (2001). Voice source characteristics in mongolian throat singing studied with high-speed imaging technique, acoustic spectra, and inverse filtering *Journal of Voice* 15/1: 78-85.

Hertegård S, Granqvist S, Lindestad P-Å (2000). Botulinum toxin injections for essential voice tremor. *Annals of Otology, Rhinology & Laryngology.* 109/2: 204-209.

Hertegård S, Larsson H, Granqvist S (2000). Realistic measures of dimensions and calibration of laryngeal images, oto-rhino-laryngology, head and neck surgery. Berlin, May 13-18.

Gauffin J, Granqvist S, Hammarberg B, Hertegård S, Håkansson A (1995). Irregularities in the voice: some perceptual experiments using synthetic voices. *Proc. of ICPhS-95* 2: 242-245.

Granqvist S (1996). Addsynt, an additive voice synthesiser for PC. *TMH-QPSR, KTH,* 4/1996: 57-60.

Granqvist S, Gauffin J, Hammarberg B, Hertegård S (1997). Perceptual sensitivity for spectral position of subharmonics in synthesised pathological voices. *Poster presentation PEVOC II* Regensburg 1997.

Other

Dahl S, Granqvist S, Thomasson M (2000). Detection of drift in tempo TMH-QPSR, KTH, 4/2000: 19-27.

Dahl S, Granqvist S (2002). A new method for estimating internal drift and just noticeable difference in perception of continuous tempo drift. *Poster presentation.* The Neurosciences and Music, Mutual Interactions and Implications on Developmental Functions. Fondazione Pierfranco e Luisa Mariani, Venice, San Servolo Island, 25-27 October, 2002.

Gleiser J, Friberg A, Granqvist S (1998). A method for extracting vibrato parameters applied to violin. *TMH-QPSR, KTH*, 4/1998: 39-44.

Appendix II

Computer programs

Some of my computer programs have been included in the Soundswell package. These include *Audiofil*, *Extract, Resample, Beep, Swellcal* and also *Corr* which produce the correlogram that was described in paper II. The program *Visor* and *Judge* that was examined in paper I is part of the Spruce listening test package together with *Glue*. I also wrote the user interface of the phonetogram program *Phog* in association with Jonas Engdegård who wrote the DSP parts. All these programs are made available by Hitech Development AB.

AddSynt, A voice synthesiser built on additive synthesis. This computer program was described in TMH-QPSR 4/96, 57-60, *Aura* is the program used for the binaural technique in paper III, *LarVib* produce the Fourier analysis of high-speed laryngoscopy as presented in paper IV, *Decap* is a program for manual inverse filtering of microphone or airflow signals that was used in paper V. *Beat* was developed in cooperation with Sofia Dahl, and the main purpose is to examine the perception of tempo drift in a sequence of clicks. *Tombstone* is a replacement for the old B&K sweep measurement equipment and is used for measurement of transfer functions. *RTSect* is a real-time spectrum analyser. *Tone* is a tone generator replacement, *Madde* is a real time singing synthesiser, in the spirit of the Musse synthesiser developed at TMH. However, it uses additive synthesis, which is advantageous when spectral properties of the voice quality are of interest.

Although these programs do not all represent major scientific progress, some of them have become widely used as tools in research, and also for pedagogical purposes.

References

These are all references that occur in the papers and the introduction.

Airo E, Olkinuora P, Sala E (2000). A method to measure speaking time and speech sound pressure level. *Folia Phoniatr Logop* 52: 275-288.

Aronsson C. Evaluation of an automatic phonetogram method using speech material recorded during a working day – a relevant method for vocal loading. Unpublished, available on request from the author.

Badin P, Hertegård S, Karlsson I (1990). Notes on the Rothenberg mask, *STL-QPSR, KTH,* 1/1990, 1-7.

Baer T, Löfqvist A & McGarr N (1983). Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques *J Acoustic Soc Amer* 73: 1304-08.

Berliner J, Durlach NI, Braida LD (1978). Intensity perception. IX. Effect of a fixed standard on resolution in identification. *Journal of the Acoustic Society of America*, vol 64, no 2, 687-689.

Blomgren M, Chen Y, Ng M, Gilbert H (1998). Acoustic, aerodynamic, physiologic and perceptual properties of modal and vocal fry registers. *J Acoust Soc Amer*, 103, 2649-2658.

Buekers R, Bierens E, Kingma H, Marres EHMA (1995). Vocal load as measured by the voice accumulator. *Folia Phoniatr Logop,* 47: 252-61.

Childers D, Naik JM, Larar JN, Krishnamurty AK & Moore GP (1983). Electroglottography, speech, and ultra-high speed cinematography, in I Titze and R Scherer, eds, *Vocal Fold Physiology, Biomechanics, Acoustics and Phonatory Control*, Denver, CO: The Denver Center for The Performing Arts.

DeKrom G (1995). Some spectral correlates of pathological breathy and rough voice qualitiy for different types of vowel fragments. *J Speech Hear Res* 38: 794-811.

Drioli C (2002). A flow waveform adaptive mechanical glottal model, *TMH-QPSR* 43: 69-79. http://www.speech.kth.se/qpsr/tmh/2002/02-43-069-079.pdf.

Dunker E, Schlosshauer B (1964). Irregularities of the laryngeal vibratory pattern in healthy and hoarse persons, *Proceedings of Research Potentials in Voice Physiology*. Ed, Brewer DW. State university of New York; 151-184.

Eysholdt U, Tigges M, Wittenberg T, Proschel U (1996). Direct evaluation of high-speed recordings of vocal fold vibrations, *Folia Phoniatr Logop*;48(4):163-70.

Fant G & Sonesson B (1962). Indirect studies of glottal cycles by synchronous inverse filtering and photo-electrical glottography, *STL-QPSR* 4/1962, 1-3.

Fant G (1982). Preliminaries to analysis of the human voice source, *STL-QPSR* 4/1982, 1-27.

Flanagan JL, Landgraf LL (1967). Self-oscillating source for vocal tract synthesizers, *IEEE Trans* AU-16, March 1968, 57-64.

Fourcin A (1986). Electrolaryngographic assessment of vocal fold vibration. *J Phonetics* 14: 435-442.

Fritzell B (1996). Voice disorders and occupation. *Log Phon Vocol* 21: 7-12.

Gauffin J, Granqvist S, Hammarberg B, Hertegård S, Håkansson A (1995). Irregularities in the voice, some perceptual experiments using synthetic voices. *Proc of XIII Intl Congress of Phonetic Sciences (ICPhS95)*, Stockholm, 2: 242-245.

Gerratt B, Kreimann J, Antonanzas-Barroso N, Berke G (1993). Comparing internal and external standards in voice quality judgements. *Journal of Speech and Hearing Research*, 36: 14-20.

Gleiser J, Friberg A, Granqvist S (1998). A method for extracting vibrato parameters applied to violin performance *TMH-QPSR* 4/1998: 39-44.

Granqvist S (1996). Addsynt, an additive voice synthesiser for PC. *TMH QPSR* 4/96: 57-60.

Granqvist S, Lindestad P-Å (2001). A method of applying Fourier analysis to high-speed laryngoscopy, *J Acoustic Soc Amer* 110/6: 3193-3197

Granqvist S (2000). Computer methods for perceptual, acoustic and laryngoscopic voice analysis. Stockholm: *Licentiate thesis.* Department of Speech, Music and Hearing ISBN: 91-7283-013-1.

Hammarberg B (1986). Perceptual and acoustic analysis of dysphonia. Stockholm: *Dissertation.* Dept of Logopedics and phoniatrics, Karolinska institute, Stockholm.

Hammarberg B, Gauffin J (1995). Perceptual and acoustical characteristics of quality differences in pathological voices as related to physiological aspects. In: Fujimura O, Hirano M (eds). *Vocal Fold Physiology, Voice Quality Control*. San Diego: Singular Publishing Group; 283-303.

Hammarberg B (1995). High-speed observations of diplophonic phonation. In: Fujimura O, Hirano M (eds). *Vocal Fold Physiology, Voice Quality Control*. San Diego: Singular Publishing Group; 343-345.

Hammarberg B (2000). Voice research and clinical needs. *Folia Phoniatr Logop*. 52: 93-102.

Hertegård, S, Björck G, Manneberg G (1998). Using laser triangulation to measure vertical distance and displacement of laryngeal mucosa. *Phonoscope*. 1; No 3: 179-185. 10.

Hess W (1983). *Pitch determination of speech signals*. Springer-Verlag. ISBN 0-387-11933-7.

Hess W (1995). Determination of glottal excitation cycles in running speech. *Phonetica* 52: 196-204.

Hillenbrand J (1988). Perception of aperiodicities in synthetically generated vowels. *J Acoust Soc Am* 83: 2361-2371.

Holmberg EB, Hillman RE, Hammarberg B, Södersten M, Doyle P (2000). Efficacy of a behaviorally-based voice therapy protocol for vocal nodules. Accepted 2000 for *J.of Voice*.

House D (2000). Rise alignment in the perception of focal accent and pitch in Swedish. *Proc. Fonetik 2000*, Skövde, Sweden 73-76.

Imaizumi S (1986). Acoustic measures of roughness in pathological voice. *J Phonetics* 14, 457-462.

Ishiki N, Okamura H, Tanabe M, Morimoto M (1969). Differential diagnosis of hoarseness. *Folia Phoniatrica* 21: 9-19.

Karnell M, Scherer R, Fischer L (1991). Comparison of acoustic voice perturbation measures among three independent voice laboratories. *J Speech Hear Res* 34: 781-790.

Kiritani S, Imagawa H, Hirose H (1988). High-speed digital image recording for the observation of vocal cord vibration. In: *Vocal Physiology: Voice Production, Mechanisms and Functions.* Fujimura O (ed.). 261-269.

Kiritani S (1995). Recent advances in high-speed digital image recording of vocal cord vibration, *Proceedings of International Congress of Phonetic Sciences* 4: 62-67.

Kitzing P (1979). Glottographic frequency analysis (in Swedish). *Doctoral dissertation*. Lund University, ENT-clinic, Malmö, Sweden.

Kreimann J, Gerrat B, Berke G (1994). The multidimensional nature of pathologic voice quality. *J Acoust Soc Amer* 96: 1291-1302.

Kreimann J, Gerratt BR (1996). The perceptual structure of pathologic voice quality. *J Acoust Soc Amer* 100: 1787-1795.

Kreimann J, Gerrat B (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research* 36: 21-40.

Köster O, Marx B, Gemmar P, Hess M, Künzel HJ (1999). Qualitative and quantitative analysis of voice onset by means of multidimensional voice analysis system (MVAS) using high-speed imaging, *Journal of Voice* 13: 355-374.

Laver J (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press, Cambridge. ISBN 0-521-231760.

Ladefoged P (1988). Discussion of phonetics: a note on some terms for phonation types. In: Fujimura O (ed), *Vocal physiology: Voice production, mechanisms and functions* New York: Raven Press; 373-375.

Larsson H, Hertegård S, Lindestad P-Å, Hammarberg B (1999). Vocal fold vibrations: High-speed imaging, kymography and acoustic analysis, *TMH-QPSR, KTH*, Stockholm 1-2/1999 21-29.

Larsson H, Hertegard S, Lindestad PA, Hammarberg B (2000).Vocal fold vibrations: high-speed imaging, kymography, and acoustic analysis: a preliminary report. *Laryngoscope*. 110/12: 2117-2122.

Liljencrants J (1991). A translating and rotating mass model of the vocal folds. *STL-QPSR, KTH*, Stockholm 1/1991: 1-18.

Liljencrants J (1996). Analysis by synthesis of glottal airflow in a physical model, *TMH-QPSR, KTH,* 2/1996, 139-142.

Masuda T, Ikeda Y, Manako H, Komiyama S (1993). Analysis of vocal abuse: Fluctuations in phonation time and intensity in 4 groups of speakers. *Acta Otolaryngol* 113: 547-552.

McAllister A, Sederholm E, Ternström S, Sundberg J (1995). Perturbation and hoarseness: A pilot study of six children's voices. *Journal of Voice* 10/3: 252-261.

McAllister A, Sundberg J, Hibi S (1996). Acoustic measurements and perceptual evaluation of hoarseness in children's voices. *TMH-QPSR, KTH*, 4/1996: 15-26.

McAllister A (1997). Acoustic, perceptual and physiological studies of ten-year-old chilren's voices. *Doctoral thesis*. Dept of Logopedics and Phoniatrics, Karolinska Institute and Dept of Speech, Music and Hearing, Royal Institute of Technology (KTH), Stockholm.

Moore P, White F, von Leden H (1962). Ultra-high speed photography in laryngeal physiology, *Journal of Speech and Hearing Disorders* 27/2: 165-171.

Neils LR, Yairi E (1987). Effects of speaking in noise on vocal fatigue and vocal recovery. *Folia Phoniatr* 39: 104-112.

Neubauer J, Mergell P (2000). Extraction and analysis of spatio-temporal glottal contour patterns: high-speed glottography and nonlinear dynamics. *Proc. Advances in Quantitative Laryngoscopy, Voice and Speech Research,* Jena Germany, April 7-8, 2000.

Ohlsson A-C, Brink O, Löfqvist A (1989). A voice accumulator - validation and application. *J Speech Hear Res* 32: 451-457.

Omori K, Kojima H, Kakani R, Slavit D, Blaugrund S (1997). Acoustic characteristics of rough voice: Subharmonics. *J Voice* 11: 40-47.

Popolo P, Švec J, Rogge-Miller K, Titze I (2002). Technical considerations in the design of a wearable voice dosimeter. *Poster presentation at ASA Cancun*, Dec. 2002.
http://192.107.173.4/peter_html/Images/cancun2_112502.pdf

Pabon P (1991). Objective acoustic voice-quality parameters in the computer phonetogram. *J Voice* 5: 203-216.

Rabinov R, Kreiman J, Gerratt B, Bielamowicz S (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *J Speech Hear Res* 38: 26-32.

Rabinov R, Kreiman J, Gerratt B, Bielamowicz S (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research*, 38: 26-32.

Rantala L, Lindholm P, Vilkman E (1998). F0 change due to voice loading under laboratory and field conditions. A pilot study. *Log Phon Vocol* 23: 164-168.

Rothenberg (1973). A new inverse filtering technique for deriving the glottal airflow waveform during voicing. *J Acousic Soc Amer* 53: 1632-1645.

Rothenberg (1981). *Research Aspects of Singing*, Stockholm: Royal Sw Acad Music: Publication 33: 15-33.

Sederholm E, McAllister A, Sundberg J, Dalkvist J (1993). Perceptual analysis of child hoarseness using continuos scales. *Scandinavian Journal of Logopedics and Phoniatrics* 18: 73-82.

Sederholm E (1996). Hoarseness in ten-year old children: Perceptual characteristics, prevalence and etiology. *Doctoral thesis*. Dept of Logopedics and Phoniatrics, Karolinska Institute and Dept of Speech, Music and Hearing, Royal Institute of Technology (KTH), Stockholm.

Szabo A, Hammarberg B, Håkansson A, Södersten M (2001). A voice accumulator device: Evaluation based on studio and field recordings. *Log Phon Vocol* 26: 102-117.

Szabo A, Hammarberg B, Granqvist S, Södersten M. Methods to study pre-school teachers' voice at work - Simultaneous recordings with a voice accumulator and a DAT recorder. Accepted for publication in *Log Phon Vocol* 2003.

Sundberg J (1987). *The Science of the Singing Voice*. Northern Illinois University Press. ISBN 0-87580-120-X.

Švec J, Pešák J (1994). Vocal breaks from the modal to falsetto register. *Folia Phoniatr Logop* 46: 97-103.

Švec JG, Schutte HK. (1996). Videokymography: high-speed line scanning of vocal fold vibration, *Journal of Voice* 10: 201-205.

Švec JG (2000). On Vibration Properties of Human Vocal Folds. *Doctoral thesis*, University of Groningen, the Netherlands, ISBN: 90-367-1235-1.

Szabo A, Hammarberg B, Håkansson A, Södersten M (2001). A voice accumulator device: Evaluation based on studio and field recordings. *Log Phon Vocol* 26: 102-117.

Szabo A, Hammarberg B, Granqvist S, Södersten M. Methods to study pre-school teachers' voice at work - Simultaneous recordings with a voice accumulator and a DAT recorder. Accepted for publication in *Log Phon Vocol* 2003.

Södersten M, Granqvist S, Hammarberg B, Szabo A (2002). Vocal behaviour and vocal loading factors for preschool teachers at work studied with binaural DAT recordings. *Journal of Voice* 16/3: 356-371.

Ternström S (1994). Hearing myself with others: Sound levels in choral performance measured with separation of one's own voice from the rest of the choir. *J Voice* 8/4: 293-302.

Ternström S, Södersten M, Bohman M (2002). Cancellation of simulated environmental noise as a tool for measuring vocal performance during noise exposure. *J Voice* 16/2: 195-206.

Tigges, Wittenberg, Mergell, Eysholdt (1999). Imaging of vocal fold vibration by digital multi-plane kymography. *Computerized Medical Imaging and Graphics (CMIG).* 23/6: 323-330.

Titze I (1973). The human vocal cords: A mathematical model. Part I. *Phonetica* 28: 129-170; Part II, (1974) *Phonetica* 29: 1-21.

Titze I, Liang H (1993). Comparison of F0 extraction methods for high-precision voice perturbation measurements. *J Speech Hear Res* 36: 1120-1133.

Titze I (1994). Definitions and Nomenclature Related to Voice Quality. *Vocal Fold Physiology. Voice Quality Control,* San Diego: Singular Publishing Group; 335-342.

Titze IR, Švec JG, Popolo PS. Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *Journal of Speech, Language, and Hearing Research*. Accepted for publication.

van de Weijer J (2002). Terminal rises in infant-directed and adult-directed questions. *Proc of Fonetik 2002, TMH-QPSR, KTH*, 44: 5-8, http://www.speech.kth.se/qpsr/tmh/2002/02-44-005-008.pdf

Watanabe H, Shin T, Oda M, Fukaura J, Komiyama S (1987). Measurement of total actual speaking time in a patient with spastic dysphonia. *Folia Phoniatr* 39: 65-70.

Wewers ME, Lowe NK (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, 13: 227-236.

Wong D, Ito RI, Cox NB, Titze IR (1991). Observation of perturbation in a lumped-element model of the vocal folds with application to some pathological cases *J Acoust Soc Am*er 89: 383-394.