MEDICAL SCHOOL

Institute of Cellular Medicine

Improving jitter and shimmer measurements in normal voices

Meike Brockmann-Bauser

Thesis submitted in fulfillment of
the requirements for the degree of
Doctor of Philosophy

December 2011

Abstract

Instrumental acoustic voice analysis is a widely used clinical assessment technique to assist differential diagnosis, documentation and evaluation of treatment for voice disorders. However recent reports criticise an unsatisfactory reliability and validity of acoustic assessments. The present work examines confounding factors associated with the usual clinical measurement procedure and how their influence might be reduced. Further, it was investigated what jitter and shimmer indicate, and how this could be applied in voice clinics.

In a routine clinical voice assessment the individuals` speaking voice SPL and F_0, gender and the vowel significantly influence both jitter and shimmer. Differences in habitual voice SPL have by far the strongest influence, which may even underlie gender effects. It was shown for the first time, that clinical jitter and shimmer measurements might be considerably improved when patients phonate at a predefined level of 85dBA (10cm distance) without control of F_0, and always use the vowel /a/.

In healthy adults jitter and shimmer were not associated with perceptual voice irregularity. However clinical measurements in a variety of voice tasks showed that jitter and shimmer were always lower in higher voice intensities. Also in vocally healthy teachers increased voice SPL and F_0 after a working day were associated with lower jitter and shimmer. Higher voice SPL is accompanied by increased vocal fold tone, which might result in more regular voice vibration patterns and thereby in lower jitter and shimmer values.

From a clinical perspective this would be highly relevant information, especially in patients with impaired vocal fold tone regulation such as in functional or neurogenic voice disorders. Future research should clarify sources of jitter and shimmer and revise current normative values considering the proposed assessment protocol and gender. This might establish the clinical potential of instrumental acoustic voice analysis.

Dedication

Für Sarah

Acknowledgements

My first thank goes to Paul Carding and Michael Drinnan for being exemplary supervisors, for tiredless mentoring, helping herd my ideas and questions, and for always providing the most comprehensive feedback one could imagine. You made it a pleasure to pursue the PhD and you are great models.

Clinical research would not work without volunteers. Many thanks to my colleagues at the University Hospital Zurich as well as the teachers at several schools in Zurich whom I was able to record for this work.

Research could not develop without people giving the opportunity and background to test ideas. Plenty of thanks to KD Dr. J. Bohlender and Prof. Dr. R. Probst of the University Hospital Zurich for happily (I suppose) letting me test the newest instrumental assessment protocols and plug together assessment equipment in improbable combinations. A special thanks goes to my colleagues in the Phoniatrics and Speech Pathology Team at the University Hospital Zurich who patiently performed one novel perceptual voice analysis task after the other and pointed out many significant issues to me. Without this, my work would have been impossible.

Thanks to all patients who keep telling me what a voice problem is. I am listening.

Many, countless, thanks to Charlotte Buswell and her wonderful family for uncounted nights of B&B, plenty of great meals, always a warm welcome and wonderful discussions. In all these years your hospitality made me feel at home in Newcastle. And it made me want to come back. I wonder if I ever can ever express what it meant to me. Let’s start with: Danke, Danke, Danke.

Thanks again to Michael Drinnan for drawing me into research. I still agree with what I heard in your first lecture for my MSc course: there are many fascinating things out there. There are! My curiosity has grown ever since, and that makes it all a pleasure.

Last, but not least: Christopher, love. Thank you for your love, for believing in me all the time, for your never ending patience and support, for countless fascinating and captivating discussions. For yourself. It is a blessing to be married to you. I am grateful every day.

And Sarah, love. Just for yourself. For happiness that did not exist before.

Table of Contents

List of Tables

List of Figures

List of Appendices

Preface: Organisation of the thesis

The present thesis grew out of a need to improve our understanding of clinical instrumental acoustic voice assessments. Instrumental measurements determine defined acoustic properties of the human voice sound. It is assumed that the acoustic voice waveform represents the vibratory characteristics of the vocal folds. However, despite widespread application in voice diagnostics and documentation, the reliability and validity of instrumental assessments has not been sufficiently established to date.

This thesis presents an in-depth study of confounding factors associated with the clinical measurement procedure and focuses on the two most widely used acoustic parameters jitter and shimmer. Further, it will be investigated what jitter and shimmer might indicate and how this could be usefully applied in voice clinics. Below the organisation of the thesis is described.

Chapter 1 provides the broad theoretical context for instrumental acoustic voice assessments. First, the main vibratory principles of the human voice are described, since these are basic to a valid interpretation of jitter and shimmer. The most commonly used clinical voice examination techniques are evaluated and placed in context with acoustic assessments. *Chapter 2* gives a comprehensive introduction into the measurement technique underlying instrumental acoustic analysis. A critical literature review documents how strongly the reliability, validity and sensitivity of clinical jitter and shimmer measurements are restricted to date. Confounding factors associated with the measurement technique and the clinical voice assessment protocol are evaluated. This chapter concludes with the main questions to be addressed in the present thesis.

Chapter 3 summarises the selection criteria for all study participants and the main voice analysis technique of the present work. The experimental chapters 4 to 8 follow the same broad structure: Introduction (including a specific literature review), Specific Study Aims, Methods, Results, Discussion and Conclusion.

Chapter 4 presents a summary only of the first research study. This work was conducted as MSc project prior to the present PhD work and was submitted to Newcastle University in August 2006. A publication based on the Master thesis is attached in Appendix B. This initial study showed a dramatic influence of differences in the habitual speaking voice Sound Pressure Level (voice SPL) on both jitter and shimmer, which may even underlie gender effects. In *chapter 5*, the relative effects of voice SPL, fundamental frequency (F_0), gender and vowel and their interactions were assessed. To determine how clinical measurements of jitter and shimmer might be improved, in *chapter 6* healthy adults were assessed under a range of voice tasks. Based on the results, a revised clinical instrumental acoustic assessment protocol is proposed.

Chapters 7 and 8 address the question what jitter and shimmer might indicate in healthy voices. In *chapter 7* it was assessed if jitter or shimmer correlate with perceived voice irregularity under the commonly used instrumental acoustic assessment protocol. Since no agreement was found, *chapter 8* describes how refined instrumental acoustic assessment methods, as specified in chapter 6, are applied in practice. Healthy teachers were recorded during a working day to determine if jitter and shimmer track subtle voice changes associated with voice use. The study results imply that jitter and shimmer might indirectly indicate changes in vocal fold tone associated with adaptation to voice use. From a clinical perspective this would be highly useful in the diagnostics of so-called functional voice disorders.

Chapter 9 discusses the experimental work in the context of clinical voice diagnostics (established in chapters 1 and 2). Based on this, useful jitter and shimmer applications and future research directions are developed.

This thesis includes material from papers and abstracts that have been published. *Appendix A* provides full publishing details of peer reviewed publications based on the work presented here (Medline standard). *Appendix B* is a published paper describing the experimental work summarized in chapter 4.

Appendix C to *Appendix F* provide evidence of ethical approval for all presented experimental studies. *Appendix G* is the participant questionnaire used in the experiments described in chapters 5 to 9. *Appendix H* and *Appendix I* show the oral instructions for judges and the user manual for the specific perceptual analysis software Newcastle Audio Ranking (NeAR). These were used in the studies described in chapters 8 and 9. *Appendix J* is the visual analogue scale for subjective voice assessment as used in chapter 9.

Chapter 1. Vibratory characteristics of voice production

Instrumental acoustic analysis of the human voice is a commonly used assessment technique. Basic to a valid interpretation of instrumental acoustic measurements is a thorough understanding of undisturbed and pathologic voice function. This chapter introduces into the main principles of human voice production. Basic anatomy and physiology as well as the vibratory properties of the vocal folds are explained. The most commonly used clinical voice assessment methods are described and evaluated with emphasis on their diagnostic properties. In this chapter instrumental acoustic assessments are placed into context of the main clinical examination techniques.

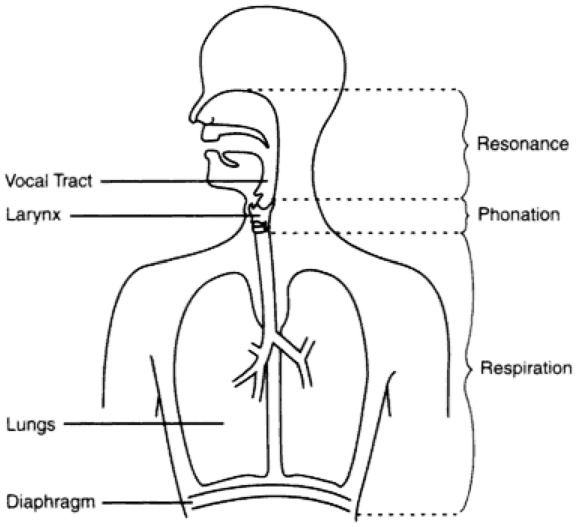
1.1 Physiology of phonation: the basis for clinical voice diagnostics

Clinical voice assessments aim to identify the reasons for pathological voice changes in a patient. Based on this the clinician plans an appropriate individual therapy approach. Therefore, the physiologic processes underlying the human voice production are key to understanding voice assessment techniques.

The human voice is generated within the larynx by vibration of the vocal folds and depends on an adequate airflow from the lungs (Titze, 1988). This requires a complex coordination of phonatory and respiratory muscles (Colton et al., 2006). The larynx is composed of muscles, ligaments, mucosa and cartilage, and is attached by ligaments and muscles only at the front of the neck (Figure 1). It is situated between the hyoid bone and the upper end of the trachea (Sobotta, 2006).

The unformed voice sound emitted from the vocal folds is amplified and filtered by the vocal tract (Fant, 1980). Movements of the speech organs in the vocal tract (for example the tongue) actively form the sound to create speech. In the following sections the main principles of human voice production will be described.

Figure 1: Location of the larynx



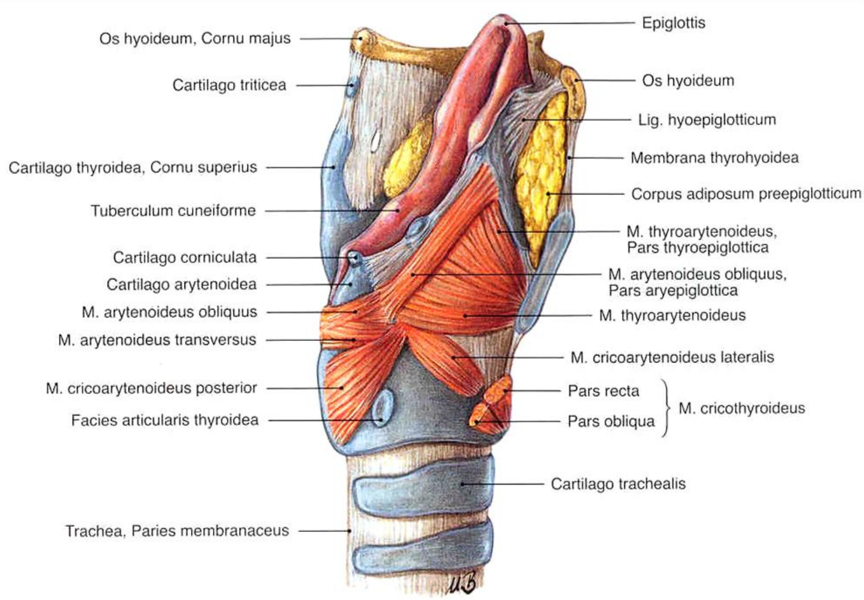
Location of the sound source, the larynx, within the human body. Indicated are the three subsystems of voice production: respiration, phonation and resonance. Origin: figure 2-1; Stemple et.al.2000.

1.1.1 *Inside the larynx: muscle activity and Bernoulli Effect*

In order to produce sound the vocal folds have to close with appropriate tone (Colton et al., 2006). This is done by three paired and one single muscle called the “adductors”: the interarytenoid muscle (“m. interarytaenoideus”), lateral cricoarytenoid muscle (“m. cricoarytaenoideus lateralis”), cricothyroid muscle (“m. cricothyroideus”), and the thyroarytenoid muscle (“m. thyroarytaenoideus” or “m. vocalis”) which forms part of the vocal folds (Figure 2). The paired muscles close the glottis (the space between the vocal folds) by active movement, whereas the m.vocalis produces intrinsic tension. The only vocal fold abductor (“opener”) is the posterior cricoarytenoid muscle (“m. cricoarytaenoideus posterior”).

When the vocal folds are in phonation position (closed glottis), the respiratory system has to provide a constant airflow. The closed vocal folds provide resistance to exhaled air from the lungs (Titze, 1988). When the air bursts through the vocal cords, the pressure between the vocal folds drops. This leads to underpressure sucking the vocal folds back together, a phenomenon known as the *Bernoulli Effect*. This happens repeatedly and results in vocal fold vibration, perceived as a natural human voice sound (Baken and Orlikoff, 2000c).

Figure 2: Side view of laryngeal structures and muscles



Laryngeal cartilages and muscles from the right side. The right part of the thyroid cartilage is removed. Origin: figure 225; Sobotta, 2006.

1.1.2 *Vocal fold structure and its role in phonation*

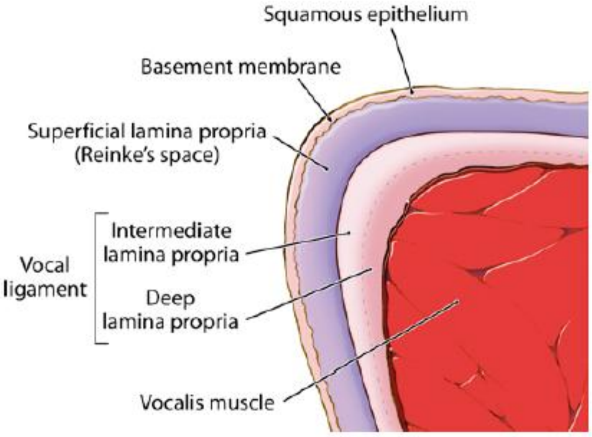
Alongside an adequate closure and tension the biomechanical structure of the vocal folds influence the vibration patterns. The vocal folds are composed of three main functional layers, the cover, the transition and the body (Figure 3). These comprise histologically different cell types with unequal vibratory properties (Table 1). As the histological layers change from superior to inferior, each level gradually changes in mass, stiffness and compliance for vibration (Hirano, 1974):

(1) The cover layer consists of *mucous membrane* and *squamous epithelium* and is very pliable.

(2) Immediately beneath lies the *lamina propria* composed of three sections. The superficial layer is called *Reinke's space* and consists of loose fibres and matrix (connective tissue). It has a jelly-like texture and easily slides over the deeper structures. Below are the *intermediate layer*, elastic fibres almost parallel to the vocal folds, and the *deep layer*, mostly collagenous fibres. Both the intermediate and the deep layer form the *vocal ligament*, which are more compact and less flexible than the superior layers.

(3) The deepest structure of the vocal folds is the *vocalis muscle (thyroarytenoid muscle)*, mainly composed of muscle fibres. This layer is rather stiff in consistency.

Figure 3: Transverse section of the vocal folds



Coronal transverse section through the vocal folds showing their histologic and functional layers. Origin: figure 1.4; Rosen et al., 2008.

Table 1: Anatomical and functional sections of the vocal folds

Anatomic layer	**Cell type**	**Function**	**Body-cover theory**	**Flexibility**
Epithelium				
	Squamous epithelium	Mucosa	Cover	Very pliable
Lamina propria				
Superficial layer (Reinke`s space)	Loose fibres, matrix	Mucosa	Cover	Very flexible
Intermediate layer	Elastic fibres	Vocal Ligament	Transition	Less flexible
Deep layer	Collagenous fibres	Vocal Ligament	Transition	Less flexible
Vocalis muscle (m. vocalis)				
	Muscle fibres	Muscle	Body	Stiff

Summary of histological and functional vocal fold layers and their vibratory properties. Origin: own table.

1.1.3 *Body-cover theory of vocal fold vibration*

Based on the vocal fold structure described above Hirano proposed a model to explain vocal fold vibration, the *body-cover theory* (Hirano, 1981). According to his theory, the cover (including the mucosa and the superficial lamina propria layer) is the main vibratory component of the vocal folds. The cover shows a compliant fluid-like oscillation termed “mucosal wave”, which can be seen in videolaryngostroboscopic examination of the larynx (section 1.2.2). Corresponding to the human voice range, the vibration rate of the cover shift (mucosa) is around 100- 1000Hz. The *transition*, the vocal ligament, serves as coupling between the flexible cover and the rather stiff *body* below, the vocalis muscle. The role of the vocalis muscle is to provide stability and an appropriate tone.

The mucosal wave is a key characteristic to assess the vibratory properties of the mucosa and the superficial lamina propria layer. Excessive, irregular and asymmetric mucosal waves appear in patients with reduced vocal fold tone (Hirano, 1981). The absence of a mucosal wave has been associated with pathology in the cover (the mucosa) or with an inadequate tension in the body (the vocal folds) (Casiano et al., 1992).

1.1.4 *Control of voice pitch and loudness*

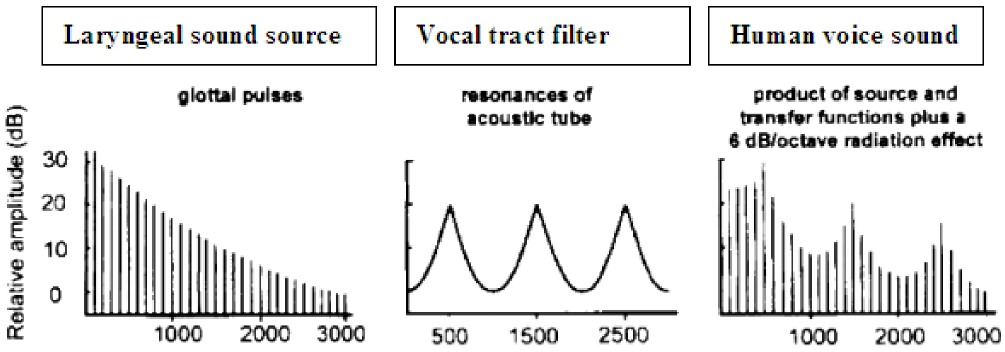
A higher voice pitch or fundamental frequency (F_0) is produced by coordinated contraction of the m.vocalis and the m.cricothyroideus. Thereby, the vocal folds are indirectly elongated and the mucosa cover stiffens. This causes the mucosa to vibrate faster, which results in a higher voice. A lower voice pitch is generated by contraction of the m.vocalis only. This shortens the vocal fold length and reduces tension in the mucosa cover, leading to lower mucosa vibration rates (Colton et al., 2006; Hirano, 1981).

Vocal intensity or loudness, as physically measured in deciBel (dB), is controlled by glottal closure. A stronger vocal fold adduction associated with increased vocal fold tone leads to higher glottal resistance. The subglottal air pressure increases until it is sufficiently high to overcome the resistance level. Thus, a more intense vibration of the vocal folds is triggered, which leads to a louder acoustic sound (Colton et al., 2006).

1.1.5 *Human voice sound modification and speech production*

According to the *Fourier-Theorem*, the natural sound signal from the sound source, the larynx, consists of a fundamental frequency (F_0) and its harmonics (Baken and Orlikoff, 2000c). The *source-filter model* describes how this sound is passively filtered through the vocal tract and actively altered by articulation to create speech (Fant, 1980).

Figure 4: Filter function of the vocal tract for the human voice



Schema of the vocal tract filter and amplification function. The larynx produces an unformed natural sound by vocal fold vibration, which is altered by the resonant properties of the vocal tract. Origin: modified image from figure 7-37; Baken and Orlikoff, 2000c.

Due to its own resonant properties the vocal tract acoustically *damps* and *amplifies* specific frequencies of the overtone spectrum of the natural sound produced by the larynx (Figure 4). In physical terms the vocal tract behaves like a composition of resonators to the sound signal. This resonance effect is termed “filter function” or “transfer function” to the original sound signal and produces spectral sound characteristics, that are specific for an individual human voice (Baken and Orlikoff, 2000c).

During speech, the size, shape and constriction of the vocal tract is voluntarily changed by the articulators: the lips, the jaw, the tongue and the velum (Ladefoged and Maddieson, 1996). This leads to a change of the resonant frequencies of the vocal tract, resulting in a different acoustic product. Thus, different language sounds are produced, which are identified as phonemes or speech by a listener (Baken and Orlikoff, 2000c).

1.2 Clinical voice assessment

1.2.1 *Voice pathology- why patients attend a voice clinic*

A comprehensive literature review by Schwartz and Cohen et al. reports, that patients usually seek professional help when their voice quality, performance or subjective vocal effort are altered (Schwartz, Cohen et al. 2009; Van Houtte, Van Lierde et al. 2010). Voice disorders may result from changes in the structure (such as vocal fold nodules or tumours), innervation (such as laryngeal nerve palsy) or function of the voice production mechanism (such as in muscle tension dysphonia) (Van Houtte, Van Lierde et al. 2010).

The likelihood to experience a voice disorder in the course of one's lifetime lies between 29% and 46% for any adult (Roy, Merrill et al. 2004; Schwartz, Cohen et al. 2009). In dysphonic adults, social participation and activity are significantly reduced (Schwartz, Cohen et al. 2009; Kleemola, Helminen et al. 2011). Further, quality of life significantly decreases with the severity of dysphonia (Jones, Carding et al. 2006). In professional voice users (such as teachers or call-centre agents), who comprise around 17% of the population, a voice disorder causes absence from work and reduces productivity (Schwartz, Cohen et al. 2009). A professional voice user may even be forced to change profession when the voice disorder is persistent (Van Houtte, Claeys et al. 2010). In summary, voice disorders can have huge effects on a patient`s personal and professional life. Therefore, efficient voice diagnostics and treatment are key to avoiding long-term negative consequences for the patient and to minimise the associated health care costs (Van Houtte, Claeys et al. 2010).

1.2.2 *Voice assessment guidelines*

For a comprehensive voice assessment European clinician associations and also voice research groups recommend a multidimensional examination of all aspects of voice production (Carding et al., 2009; Mehta and Hillman, 2008; Ma and Yiu, 2006; Dejonckere et al., 2001). The main assessment techniques include:

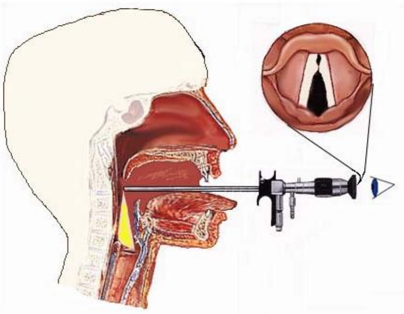
(a) Visual examination of the laryngeal structures such as by videolaryngoendoscopy (Dejonckere et al., 2001)
(b) Standardised auditory perceptual rating of voice quality such as by the Grade-Roughness-Breathiness-Asthenia-Strain (GRBAS) rating scale (Hirano, 1981)
(c) Patient self-reporting of vocal symptoms by questionnaire such as the Voice Handicap Index (VHI) (Jacobson et al., 1996)
(d) An assessment of vocal performance such as by voice range profile measurements (Heylen et al., 2002)
(e) Aerodynamic measurements such as maximum phonation time (Kent et al., 1987)
(f) Indirect measurements of laryngeal opening and closing such as by electroglottography (Baken and Orlikoff, 2000d)
(g) Instrumental acoustic assessment of the voice sound by measuring fundamental frequency (F_0), voice SPL and voice perturbation (mostly jitter and shimmer) (Baken and Orlikoff, 2000c)

In the following sections the main techniques and their clinical applications will be reviewed.

1.2.3 *Visual examination of the larynx*

Visualisation of the vocal fold structure and function is considered the standard method to diagnose vocal fold pathology (Deliyski and Hillman, 2010; Mehta and Hillman, 2008). Visual laryngeal examination techniques such as (video)laryngoendostroboscopy employ an endoscope (and a camera system) to inspect the larynx and the neighbouring structures under a range of phonatory tasks (Figure 5). Usually, the morphology of the tissue, the quality of vocal fold movement and closure, and the vibration of the mucosa are evaluated (section 1.1.3, Dejonckere et al., 2001; Hirano, 1981).

Figure 5: Schematic image of laryngoscopic examination



Schematic overview of visual laryngeal examination by endoscopy. In this example the examiner uses a rigid endoscope to assess laryngeal morphology. Origin: website “Dysphonia”, figure title “Indirect laryngoscopy using rigid endoscope”; Akerlund et al. 2009.

Depending on the size, laryngeal pathology can be reliably diagnosed by laryngoendoscopic examination techniques. However, very small pathological lesions (Mehta et al., 2010) or functional voice disorders associated with pathological muscle tension in the vocal folds might not be recognised (Schneider et al., 2002). Further, between 36% and 50% of vocally healthy adults show signs of an abnormal vocal technique in videolaryngoscopic assessments (Sama et al., 2001). Therefore, especially in early voice pathology or in functional voice disorders the correct diagnosis relies on the assessment of instrumental acoustic, perceptual and subjective voice symptoms.

1.2.4 *Auditory perceptual analysis*

Perceptual analysis of the human voice is based on an auditory assessment with defined criteria. An expert listener rates to what extent a pathological voice deviates from a perceived “normal” level in each perceptual parameter. One of the most widely used assessment schemes is the Grade-Roughness-Breathiness-Asthenia-Strain (GRBAS) rating scale (Carding et al., 2009; Hirano, 1981).

Perceptual voice analysis has high validity, a good sensitivity to voice change and an excellent clinical utility (Carding et al., 2009). Also, a better reliability as compared to instrumental acoustic voice assessments was reported (Carding et al., 2009). However, the interrater reliability has been characterized as unsatisfactory (Mehta and Hillman, 2008; Kreiman and Geratt, 2007; Kreiman and Geratt, 2000). Therefore, perceptual assessments have limited comparability between examiners or between clinics (Kreiman and Geratt, 2007).

1.2.5 *Patient based self-reporting of vocal symptoms*

Clinical voice questionnaires such as the Voice Handicap Index (VHI) assess the patient`s subjective functional, emotional and physical voice symptoms (Jacobson et al., 1996). Subjective voice assessment by questionnaire has a high pragmatic utility, high validity and good sensitivity to change (Carding et al., 2009). Further, questionnaires indicate fluctuations in voice disorders, that might not be observed in clinical assessments (Jones et al., 2006). Subjective voice symptoms might not be associated with a visible (by laryngoscopy) or audible (by perceptual assessment) voice disorder. Since questionnaires reflect the patient`s satisfaction with their voice, they are key to determine treatment success (Carding et al., 2009).

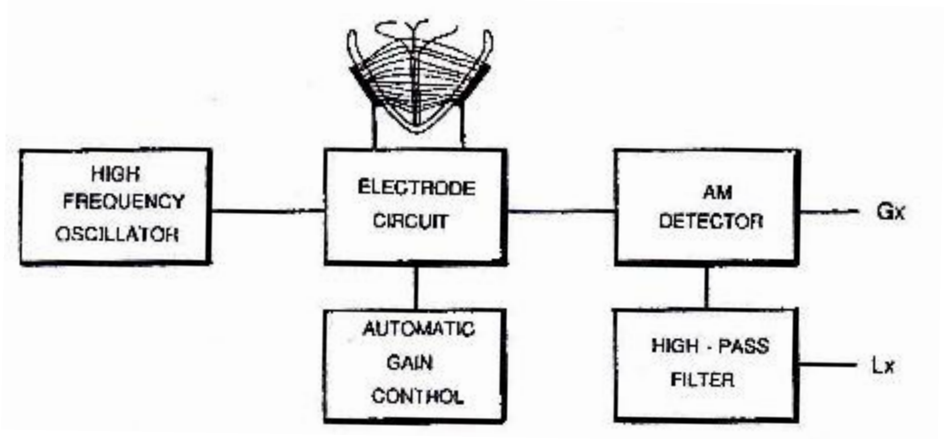
1.2.6 *Electroglottography*

Electroglottography (EGG) is used to indirectly assess vocal fold opening and closing. Figure 6 displays a schema of a typical electroglottograph. For electroglottography, two small electrodes are attached at the front of the neck, superficial to the thyroid cartilage of the larynx (Figure 1 and Figure 2). Between the electrodes passes a small high-frequency alternating current of >300KHz and <10mA. To the current the neck tissue acts as a volume conductor (Baken and Orlikoff, 2000d). According to Ohm´s law an electrical pressure difference (measured in Volt) is created across a conductor, when an alternating current interacts with impedance (Baken, 1992). The voltage increases proportionally to the rise in impedance. Neck tissue impedance is influenced by glottal opening and closing: it drops when the glottis closes, and increases when the glottis opens (Baken, 1992).

Figure 7 shows details of simultaneously recorded acoustic and electroglottographic waveforms of a healthy adult, as displayed by the voice analysis software PRAAT (Boersma and Weenink, 2006). The peaks of the raw EGG waveform indicate complete glottal closure, when the impedance and voltage are lowest and the current is at its maximum. However, transneck impedance (and with it the EGG signal) is also influenced by the position of the larynx and the head, as well as the neck shape and structure. Therefore, the reliability of electroglottographic measurements has been criticised as limited (Baken and Orlikoff, 2000d).

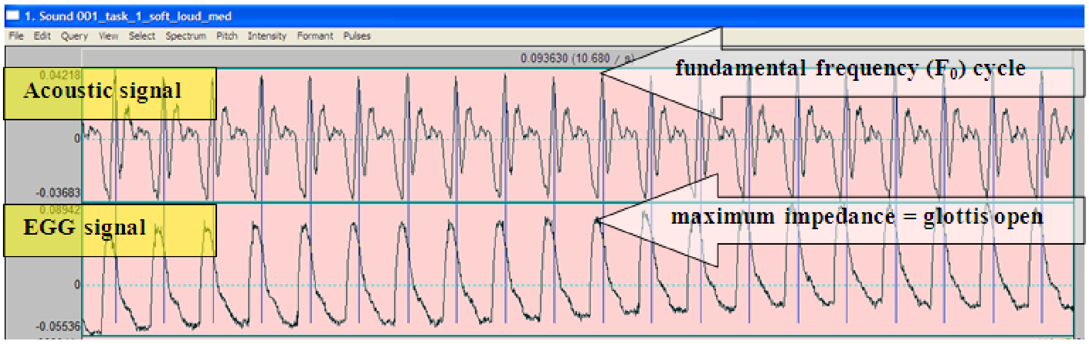
Electroglottographic measurements are mostly used to indirectly assess the degree of vocal fold closure, for example in patients with laryngeal palsy (Zagólski, 2009; Colton and Conture, 1990; Haji et al., 1986). However, electroglottography is not routinely applied in all voice patients (Dejonckere et al., 2001).

Figure 6: Representative schema of a typical electroglottograph



Schematic overview for a typical electroglottograph as used in voice clinics. Origin: figure 2, Baken, 1992.

Figure 7: Simultaneously recorded acoustic and electroglottographic signal



Screenshot of simultaneous acoustic and electroglottographic voice recording as displayed by the software PRAAT (Boersma and Weenink, 2006). The upper window section shows the acoustic signal, and the lower section the electroglottographic wave. Vertical blue lines indicate the fundamental frequency cycles in the acoustic voice signal. Acoustic waveform and EGG cycles are not fully symmetrical. Origin: own image.

1.2.7 *Instrumental acoustic voice assessments*

Instrumental acoustic measurements are used to assess defined physical properties of the acoustic voice signal. In clinical voice diagnostics mostly measurements of voice range profiles and of voice perturbation are used (Dejonckere et al., 2001). Voice range profiles or phonetograms consist of fundamental frequency (F_0) and voice Sound Pressure Level (voice SPL) measurements under a range of voice tasks (Pabon, 1991). Individual profiles of the mean, minimum and maximum F_0 and voice SPL in speaking and singing voice tasks are considered to reflect the patient's vocal capacity (Dejonckere et al., 2001).

The present work focuses on voice perturbation measurements, which usually include measurements of jitter (F_0 variation from one acoustic wave to the next) and shimmer (SPL variation) (Brockmann-Bauser and Drinnan, 2011; Carding et al., 2004). Voice perturbation measurements are based on the idea that the acoustic waveform represents the vibratory characteristics of the vocal folds. Increased jitter and shimmer have been associated with vocal pathology and dysphonia (Mehta and Hillman, 2008). Also, it has been suggested that jitter and shimmer might indicate vocal pathology not perceptible by the human ear (Stojadinovic et al., 2002). Please refer to chapter 2 for detailed description of these parameters and the underlying measurement technique.

Despite the widespread and routine use, the reliability and validity of acoustic jitter and shimmer has been criticized as not satisfactory for a useful clinical application (Brockmann-Bauser and Drinnan, 2011; Carding et al., 2004). In the following chapter the evidence base underlying the clinical applications of jitter and shimmer will be evaluated in detail (chapter 2).

1.3 Summary

Clinical voice assessments aim to identify the reasons for pathological voice changes in a patient. Therefore, an understanding of the anatomy and physiologic processes underlying the human voice are key for a successful voice diagnostics.

The human voice is generated by vocal fold vibration within the larynx. This requires a complex coordination of phonatory and respiratory muscles. Vocal fold vibration patterns are mainly influenced by the degree of vocal fold closure and the status of the covering mucosa. Voice pitch and intensity are mainly modulated by intrinsic tension and closure of the vocal folds. Further, the voice signal is altered by vocal tract filtering and articulation to create speech.

A comprehensive clinical voice examination usually includes visual, perceptual, patient based subjective and instrumental acoustic assessment techniques. Currently it is not possible to reliably detect small or early pathological changes in the vocal folds by videolaryngostroboscopy. Also, disorders associated with discrete pathology in muscle tone regulation such as muscle tension dysphonia (or functional dysphonia) are difficult to diagnose by videolaryngostroboscopy. In these cases the diagnosis relies on perceptual and instrumental acoustic and subjective voice assessments.

Despite a widespread and routine use the reliability and validity of acoustic jitter and shimmer has been criticized as not satisfactory for a useful clinical application. The evidence base underlying the clinical application of jitter and shimmer will be evaluated in detail in the following chapter 2.

Chapter 2. Principles of instrumental acoustic analysis of the human voice

Sections 2.3 and 2.4 of this chapter provided the main material for the following published work:

Brockmann-Bauser, M. and Drinnan, M. J. (2011) 'Routine acoustic voice analysis: time to think again?', Current Opinion in Otolaryngology, Head and Neck Surgery, 19, (3), pp. 165-70.

This thesis focuses on the most widely used instrumental acoustic parameters "jitter" and "shimmer". Chapter 2 provides a comprehensive introduction into the definitions of jitter and shimmer, the underlying measurement and analysis technique and the standard clinical assessment procedure. Current technical and clinical acoustic assessment guidelines are summarised. On the basis of an extensive literature review the reliability and validity of jitter and shimmer measurements are critically discussed. Confounding factors associated with the measurement technique (such as background noise) and the clinical voice task (such as differences in speaking voice intensity) are described in detail. Later chapters refer to the literature summarised in this chapter. Therefore, broad topics are subsumed in sections, which finish with a short summary of the main issues. This chapter concludes with a summary of the main aims of the present thesis.

2.1 Introduction

2.1.1 *What is acoustic voice analysis?*

Instrumental acoustic analysis of the human voice is a common assessment technique used in the study of voice pathology (Carding et al., 2009; Carding et al., 2004; Dejonckere et al., 2003; Titze, 1995). In general, instrumental acoustic voice analysis refers to a family of computer-based techniques which measure defined acoustic signal properties of a spoken (prolonged) vowel or speech (Carding et al., 2004; Titze, 1995). These include measurements of the human voice pitch (fundamental frequency, F_0) and loudness (amplitude, voice SPL), frequency and amplitude perturbation indices (such as jitter and shimmer), estimates of the proportion of aperiodicity (as signal-to-noise-ratio), spectral analysis based techniques including cepstral analysis, and methods based on nonlinear dynamics and chaos analysis (Maccallum et al., 2010; Carding et al., 2009; Mehta and Hillman, 2008; Dejonckere et al., 2001; Titze, 1995).

In studies involving clinical voice patients, the instrumental acoustic parameters jitter and shimmer are by far the most commonly applied acoustic voice measures. A representative Medline search of the past two years identifies 86 papers using these parameters. The next most common measures are “*signal-to-noise ratio*” and “*nonlinear dynamics*” with 16 entries each. Therefore, the present work focuses on clinical measurements of jitter and shimmer (Dejonckere et al., 2001; Titze, 1995).

2.1.2 *Definition of terms*

The instrumental parameters jitter and shimmer are calculated on the basis of fundamental frequency (F_0) and voice Sound Pressure Level (voice SPL) respectively.

Fundamental frequency (F_0), measures the human voice pitch by indicating the number of acoustic cycles per seconds in Hertz (Hz).

Voice Sound Pressure Level (voice SPL) describes the loudness or amplitude of the voice in deciBel (dB).

Figure 8 shows a schematic image of two acoustic waves: the red wave illustrates a sound with a lower fundamental frequency (F_0) and smaller voice Sound Pressure Level (voice SPL). This is perceived by the human ear as a deeper and softer voice sound.

Figure 8: Schema of acoustic signals with different F_0 and voice SPL

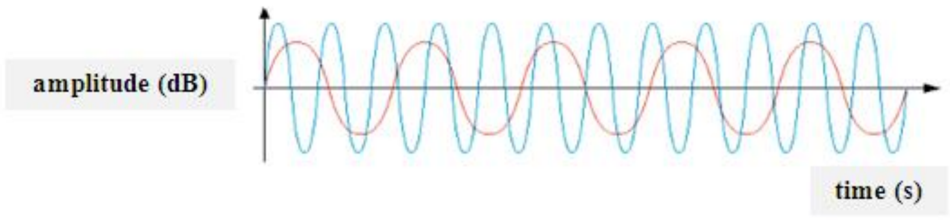
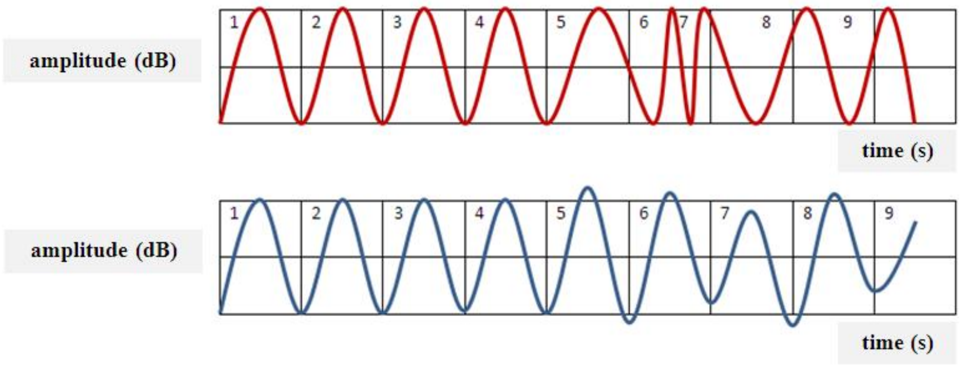


Illustration of two acoustic waves (sine waves) with a different amplitude (voice loudness) and fundamental frequency (pitch). A fundamental frequency (F_0) is the number of acoustic cycles per second measured in Hertz (Hz). In this example the blue sine wave shows a higher F_0 (pitch) and voice SPL (voice loudness) than the red wave. Origin: own figure.

Jitter measures F_0 variation from one acoustic wave to the next. Analogously, the index *shimmer* quantifies voice SPL variation from one acoustic cycle to the next.

Figure 9 illustrates the effects of fundamental frequency (jitter) and voice SPL (shimmer) irregularity on a sine wave.

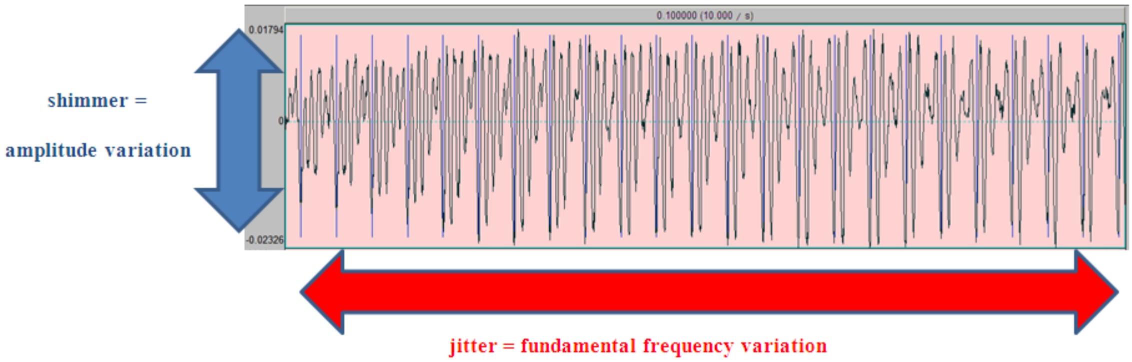
Figure 9: Effects of jitter and shimmer on sine waves



Effects of jitter (red) and shimmer (blue) on sine waves. In both examples the first 4 acoustic cycles have a regular fundamental frequency (F_0) and amplitude (voice SPL). Cycles 5 to 9 show fundamental frequency (red = jitter) and amplitude (blue = shimmer) perturbation respectively. Origin: own figure.

Figure 10 shows an oscillogram, a display of the acoustic waves, of a voice recording. In this example of a healthy adult voice, visible jitter (fundamental frequency irregularity) and shimmer (amplitude irregularity) are present.

Figure 10: Oscillogram of acoustic human voice signal



Oscillogram of a normal voice as displayed by the instrumental acoustic analysis software PRAAT. In this voice example visible shimmer (amplitude variation) and jitter (fundamental frequency variation) are present. Origin: own figure.

2.1.3 *Acoustic or electroglottographic jitter and shimmer?*

Instrumental parameters such as jitter and shimmer may be derived from the acoustic or the electroglottographic waveform (please see section 1.2.5). Electroglottographic jitter and shimmer have been characterised as more sensitive than acoustic voice perturbation. Further, Orlikoff stated that electroglottographic jitter and shimmer may be uninfluenced by vocal tract effects (Baken and Orlikoff, 2000d; Orlikoff, 1995). However, research describing the agreement between acoustic and electroglottographic jitter and shimmer is still limited (Baken and Orlikoff, 2000d). In voice clinics almost exclusively acoustic jitter and shimmer measures are used (Brockmann-Bauser and Drinnan, 2011; Carding et al., 2009; Baken and Orlikoff, 2000a; Baken and Orlikoff, 2000b).

It has been hypothesised that electroglottgraphic jitter indicates frequency stability, and therefore is considered to be similar to acoustic jitter (Baken and Orlikoff, 2000d). However, as shown by Figure 7, the maximum closure of the vocal folds as indicated by the maximum voltage (lowest impedance) does not exactly agree with the fundamental frequency cycle in the acoustic signal. Therefore, EGG jitter might be systematically different from acoustic jitter. In turn, EGG shimmer has been described as a completely different measure than acoustic shimmer. Whereas acoustic shimmer measures differences in the acoustic wave amplitude, EGG shimmer might indicate changes associated to the degree of vocal fold closure (Baken and Orlikoff, 2000d). Based on the limited present literature it can be concluded that electroglottographic jitter and shimmer have not been satisfactorily described for a useful application in voice clinics yet (Baken and Orlikoff, 2000d).

2.1.4 *Hypothesis underpinning the interpretation of jitter and shimmer*

Both jitter and shimmer measure the irregularity (or unintentional short-term variation) of the acoustic signal produced by vocal fold vibration (Baken and Orlikoff, 2000b; Baken and Orlikoff, 2000a). The interpretation of jitter and shimmer is based on the hypothesis that the acoustic waveform represents the vibratory characteristics of the vocal folds. Therefore, it has been suggested that jitter and shimmer indirectly describe the biomechanical status of the vocal folds (Maryn et al., 2009; Mehta and Hillman, 2008) and the stability of vocal fold vibration (Baken and Orlikoff, 2000b; Fukazawa et al., 1988). It has been also presumed that jitter and shimmer might indicate different aspects of perceptual dysphonia, and thereby objectively measure the vocal output of the larynx (Ma and Yiu, 2006; Dejonckere et al., 1996).

Small irregularities in the acoustic voice signal are considered as normal variation related to physiologic body functions such as heartbeat or changes in vocal fold tension. Increased levels have been described as an indicator of vocal pathology (Titze, 1991; Orlikoff and Baken, 1989). Therefore, voice perturbation analysis has been characterised as an easily applicable, indirect and non-invasive assessment tool of vocal function.

2.1.5 *Physiologic sources of jitter and shimmer*

In clinical assessments we are confronted with natural variability between patients due to natural physiologic differences. A number of physiologic sources of variability in instrumental acoustic measurements have been described in the literature, including:

(a) aging effects (Stathopoulos et al., 2011; Sussmann and Sapienza, 1994; Linville and Korabic, 1987; Wilcox and Horii, 1980)
(b) the heartbeat (Titze, 1991; Orlikoff and Baken, 1989)
(c) aerodynamic turbulences in the vocal tract during phonation (Titze, 1988)
(d) activity of the motor neurons innervating the laryngeal muscles (Titze, 1991)

Age effects might be addressed in clinical measurements of jitter and shimmer by using age specific normative values (Stathopoulos et al., 2011). However, as criticised by Baken and Orlikoff, to date these are not available in the literature (Baken and Orlikoff, 2000c). In a study by Wilcox and Horii, 20 young men (mean age 23.3) were compared with 20 older men (mean age 69.8 years). Higher jitter was found in the older men (Wilcox and Horii, 1980). However, a meta-analysis of five studies with a total number of 51 adults between 21 and 80 years of age found, that jitter and shimmer tend to *gradually* increase with age (Baken and Orlikoff, 2000c). Further, Ramig and Ringel in a study of 48 men between 25 and 75 years, documented that irrespective of the age jitter and shimmer were lowest in men with a good physical condition (Ramig and Ringel, 1983). This shows that further investigations with an appropriate study sample are necessary to determine clinically useful comparison data.

Theoretically, perturbation measurements might be corrected for changes in the activity of motor neurons or the heartbeat rate. However, for this we would need a reliable model of how these factors influence jitter and shimmer. Again, the research base is limited and a clinically useful model to correct for these effects is not available to date (Baken and Orlikoff, 2000b). Therefore, the influencing factors described have to be considered as inherent natural variability between normal voices in clinical measurements of jitter and shimmer.

2.1.6 *Role and present applications of jitter and shimmer measurements*

Clinical acoustic jitter and shimmer measurements are used to provide supplementary information alongside visual laryngeal examination details and auditory perceptual voice assessment (Dejonckere et al., 2001; Titze, 1995). Further, it has been suggested that jitter and shimmer might indicate early pathology not discovered by other methods (Stojadinovic et al., 2002), changes in the biomechanical vocal fold properties or perceptual dysphonia (Brockmann-Bauser and Drinnan, 2011; Mehta and Hillman, 2008).

A review of the current literature shows that jitter and shimmer have been applied in almost every field of voice diagnostics, documentation and research to:

(a) assist differential diagnosis of voice disorders (Arjmandi et al., 2011; Carding et al., 2009; Mehta and Hillman, 2008; Dejonckere et al., 2001)
(b) measure the effectiveness of surgery (Sanuki et al., 2010; Vashani et al., 2010; Cheng et al., 2009; Hartl et al., 2009), radiation (Rovirosa et al., 2000), medical treatment (Hanson et al., 1997) or voice therapy (Ruas et al., 2010; Vashani et al., 2010; Roy et al., 2002)
(c) compare and rate the benefits of different intervention types (Van Lierde et al., 2010; Sjögren et al., 2008; Eksteen et al., 2003)
(d) corroborate the patient's perceived voice disability (Shao et al., 2010; Yelken et al., 2010)
(e) to help indicate chronic diseases such as fatigue (Cho et al., 2011)
(f) to study voice characteristics in specific professional groups such as singers (Brown et al., 2000) and teachers (Niebudek-Bogusz et al., 2007), or various ethnic groups (Dehgan et al., 2010; Kiliç et al., 2004)

2.1.7 *Summary*

Instrumental acoustic voice analysis measures defined physical properties of the human voice signal. Jitter and shimmer are derived from the pitch (fundamental frequency) and the amplitude (voice SPL). They indicate the involuntary variation in F_0 and voice SPL respectively from one acoustic wave to the next.

The interpretation of jitter and shimmer is based on the hypothesis that the acoustic waveform represents the vibratory characteristics of the vocal folds. Small irregularities in the vocal signal are considered as normal variation related to physiologic body functions. However, increased levels of perturbation have been associated with vocal pathology.

2.2 Measurement and analysis technique

2.2.1 *Standard equipment*

For acoustic voice analysis high quality recording equipment and specialised computer hardware and/or software are necessary. Acoustic analysis packages widely used in voice clinics such as “Multi-Dimensional Voice Program” (MDVP) consist of a microphone and computer hard- and software (KayPentax, 1993), or alternatively, of a microphone, a suitable sound card and specialised software such as “lingWaves Vospector” (LingCom, 2006). These hard- and software packages provide voice recording and analysis as well as patient data administration in one system.

When technical key characteristics are considered, also an individual high quality recording system with a microphone, a recorder and a laptop plus specialised free software such as “PRAAT” (Boersma and Weenink, 2006) can provide voice analysis of the same quality standard (Boersma, 2009; Titze, 1995).

2.2.2 *Acoustic wave extraction strategies*

Technically, the measurements of jitter and shimmer depend on correct recognition of F_0 and voice SPL (Mehta and Hillman, 2008; Baken and Orlikoff, 2000c; Deem et al., 1989; Titze et al., 1987). For this, software systems employ different acoustic waveform extraction strategies. The main techniques used are zero crossing, peak picking and waveform matching methods (Boersma, 2009; Maryn et al., 2009; Milenkovic, 1987).

In hoarse voices or in recordings with background noise, the waveform matching method provides the most exact F_0-extraction (Boersma, 2009; Maryn et al., 2009; Titze and Liang, 1993). Therefore, in clinical measurements this analysis technique should be used to avoid the confounding effects of additive noise on both jitter and shimmer (Boersma, 2009).

2.2.3 *Main calculation strategies of different jitter and shimmer types*

There are two main jitter and shimmer types basing on different general calculation strategies: *absolute* and *adjusted* measures.

Absolute jitter and shimmer measures such as “Perturbation Factor” or “Directional Perturbation Factor” are based on the *difference between successive periods or amplitudes* irrespective of the general F_0 and SPL. However, these measures have been shown to be affected by mean F_0 and SPL respectively (Orlikoff and Baken, 1990). Jitter tended to be smaller in higher F_0, and shimmer was smaller in higher voice SPL because the *absolute difference* between acoustic waves decreases with increasing frequency and amplitude.

To correct for this, F_0- and SPL- adjusted jitter and shimmer measures such as “jitter %”, “Jitter Ratio” or “shimmer dB” were introduced. F_0-adjusted indices are calculated as *a ratio of mean perturbation to mean waveform duration.* Analogously, SPL-adjusted shimmer is calculated by a ratio of mean amplitude to mean waveform amplitude. With this calculation method differences in F_0 and SPL are automatically corrected for (Baken and Orlikoff, 2000b; Baken and Orlikoff, 2000a). Therefore, F_0- and SPL-adjusted perturbation measures should be preferred in clinical measurements of jitter and shimmer.

2.2.4 *Technical guidelines regarding equipment and recording conditions*

Beside an adequate computer based waveform extraction and calculation approach, overall measurement reliability of jitter and shimmer also depends on characteristics of recording equipment, environment and procedure. The microphone type and placement (Deliyski et al., 2006; Winholtz and Titze, 1997b), recording technique (Gelfer and Fendel, 1995), environmental noise (Deliyski et al., 2006; Perry et al., 2000), the voice sample length (Karnell, 1991) and the number of analysed acoustic cycles (Scherer, 1995; Titze, 1995) significantly impact on both jitter and shimmer.

To improve measurement reliability and comparability between centres, the National Center for Voice and Speech (NCVS) USA and the European Laryngological Society

(ELS) published guidelines regarding the technical conditions and the measurement procedure (Deliyski et al., 2006; Dejonckere et al., 2001; Titze, 1995). These were understood as minimum standard. The main specifications are:

(a) *Microphone*: condenser microphone, head-mounted, off-axis positioned (45-90°), 3-10cm distance, minimum sensitivity of –60dB

(b) *Recording technique*: Digital Audio Tape (DAT) or digital recorder, 16-bit A/D conversion rate, sampling frequency of 20-100kHz, overall signal-noise ratio between 85-95dB

(c) *Room*: control of noise level and reverberation (ideally a sound treated quiet room).

(d) *Test utterances*: sustained vowel phonation, one high and one low vowel (for example: /i/ and /o/ or /a/), around 10 repetitions (Titze, 1995), with stable pitch and voice loudness

2.2.5 ***Standard acoustic analysis procedure***

In voice clinics both jitter and shimmer are almost exclusively measured in steady state vowels. Usually the patients are asked to produce a long and continuous vowel (mostly /a/, /o/ or /i/) at "*comfortable pitch and loudness*" (Carding et al., 2004; Dejonckere et al., 2001; Titze, 1995). This instruction is used to minimise the confounding effects of intentional modulations of F_0 and voice SPL associated with normal speech prosody (Carding et al., 2009; Dejonckere et al., 2001; Baken and Orlikoff, 2000c). In section 2.5 it will be discussed if these instructions are sufficiently rigorous to control for the influence of voice SPL and F_0.

Instrumental acoustic voice analysis in clinical settings normally consists of three consecutive steps (Carding et al., 2009; Dejonckere et al., 2001; Titze, 1995):

(1) *Recording* of the acoustic voice signal:

- prolonged vowel phonations with /a/, /o/, /u/ or /i/
- the examiner instructs the patient to phonate "at comfortable loudness and pitch"

(2) Computer *extraction* of fundamental frequency (F_0) and speaking voice SPL

(3) C*alculation* (by computer) of jitter and shimmer

After recording the voice signal (Step 1), computer analysis of voice F_0 and SPL (Step 2) as well as jitter and shimmer calculation (Step 3) are usually done automatically with specialised computer hardware and/or software (Boersma and Weenink, 2006; LingCom, 2006; KayPentax, 1993). Thereafter, the examiner receives the results and, if available, normative values for comparison.

2.2.6 *Summary*

Acoustic voice analysis consists of voice recording, waveform analysis, and calculation of jitter and shimmer. Waveform analysis and jitter/shimmer calculation are normally done automatically by computer. Since jitter and shimmer reliability depend on the recording and analysis techniques, guidelines with technical specifications have been published.

The main recommendations are to:

- use a condenser microphone with a minimum sensitivity of –60dB in lateral position (45-90°) with 3-10cm distance
- utilise digital voice recordings with a 16-bit A/D conversion and a sampling rate of 20-100kHz
- use an analysis software employing the waveform matching method and providing F_0 and SPL adjusted indices
- avoid background noise and to ensure an overall signal-noise ratio of 85-95dB of the recording system

2.3 **Validity, sensitivity and specificity of jitter and shimmer**

2.3.1 ***What should jitter and shimmer indicate?***

Clinical voice assessments are concerned with determining the cause of vocal dysfunction. Both jitter and shimmer have been described as objective measures of the biomechanical vibratory properties of the vocal folds, which are considered central to the determination of voice quality (Brockmann et al., 2008; Mehta and Hillman, 2008; Baken and Orlikoff, 2000c). Also, in measuring pitch and amplitude perturbation it has been presumed that jitter and shimmer objectively indicate different aspects of perceptual dysphonia (Martens et al., 2007; Ma and Yiu, 2006; Wolfe and Martin, 1997; Dejonckere et al., 1996; Eskenazi et al., 1990).

Despite the widespread use of jitter and shimmer in clinical and research settings, their reliability, validity and sensitivity to voice change have been criticised as unsatisfactory (Carding et al., 2009; Carding et al., 2004; Zyski et al., 1984). Also, jitter and shimmer have been described as not pathology specific (Baken and Orlikoff, 2000c). In the following sections it will be reviewed if jitter and shimmer are indeed independently useful measures of vocal dysfunction. We would expect a clinically relevant diagnostic tool to have the following properties:

(a) a relation with pathological abnormalities in the larynx
(b) a relation to the severity of dysphonia
(c) a relation to the outcome of interventions to treat dysphonia
(d) independence from other measures of vocal performance

2.3.2 ***Are jitter and shimmer indicators of vocal fold pathology?***

It has been reported for a number of years that the presence (Rosen et al., 2000) and perhaps even the extent of laryngeal pathology be associated with increased jitter and shimmer (Baken and Orlikoff, 2000c; Schoentgen, 1982; Murry and Doherty, 1980; Liebermann, 1963). Correct recognition of pathologic voices has been found to be as high as 92%, when perturbation parameters are used in combination with a mathematical pattern recognition model (Wang and Jo, 2007).

However Zyski et. al., studying voice patients with a variety of laryngeal pathologies, contradicted these results (Zyski et al., 1984). Depending on the analysis approach, 21% to 77% of all patients had jitter and shimmer within the range of the healthy voices (Zyski et al., 1984). Also, an analysis of recent research casts doubt on the relation between pathological changes in the vocal folds and increased jitter and shimmer. In 15 patients with anterior commissure synechia, jitter and shimmer were not associated with the degree of voice function impairment (Pfüetzenreiter et al., 2010). Or in 40 children with vocal nodules, jitter and shimmer were not associated with nodule size (Shah et al., 2008).

Despite considerable improvement and standardisation of the measurement technique (Titze, 1995), contradictory results are also still found in studies comparing different acoustic analysis systems in the same patient group. Jiang et al. assessed 21 healthy adults, 21 patients with nodules and 39 with polyps (Jiang et al., 2009). Using MDVP (KayPentax, 1993), patients with polyps had significantly higher jitter and shimmer than healthy adults, whereas vocal nodules had no measurable effect on jitter or shimmer. Yet an analysis by CSpeech (Milenkovic and Read, 1992) on the same voices showed no effect on shimmer for either pathology group and again significant differences for jitter between the normal and polyp group.

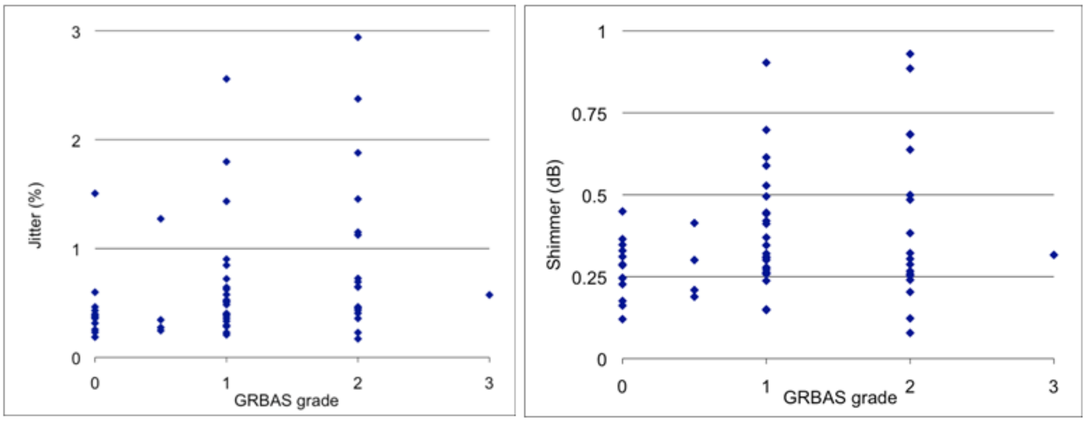
Based on these studies we must conclude that neither jitter nor shimmer has a clearly proven and unambiguous relationship with pathological vocal fold abnormalities, or with the degree of impairment to the vocal fold vibration.

2.3.3 *Are jitter and shimmer indicators of voice quality?*

Auditory-perceptual voice assessment by a standardised protocol such as the GRBAS scale (Hirano, 1981) has an established role in the evaluation of voice quality (Carding et al., 2009). Again, the link with jitter or shimmer is ambiguous. In some reports perceptual hoarseness (represented by “G” of the GRBAS scale) was accompanied by increased jitter or shimmer (Martens et al., 2007; Ma and Yiu, 2006; Dejonckere et al., 1996). Yet Bhuta et al. found no meaningful correlation for any GRBAS parameter with jitter or shimmer in 37 dysphonic patients (Bhuta et al., 2004). Similarly, Eskenazi reported no agreement of perceptual voice quality with jitter or shimmer in healthy voices (Eskenazi et al., 1990). It can be concluded that previous reports disagree on the exact nature of the relationship between GRBAS scale parameters and with jitter and shimmer.

To explain the contradictory evidence, I will now discuss experimental data on the relation of perceptual overall hoarseness with jitter and shimmer. This data has been previously published in a literature review (Brockmann-Bauser and Drinnan, 2011). Please refer to the paper for further details on the rationale and the methodology used. In this study thirty-three patients before and after radiotherapy for head-and-neck cancer were recorded while saying a prolonged /i/ for 6 seconds. GRBAS grade was rated by 2 voice experts, while jitter and shimmer were measured using PRAAT. Thereafter, the agreement between jitter and shimmer and the overall impression of "hoarseness" as indicated by "G" of the GRBAS scale was assessed (mean of two raters). This parameter was chosen since G subsumes the voice characteristics "roughness" and "breathiness", which have been previously related to jitter and shimmer. A total of 61 measurements in 33 patients with head-and-neck cancer before and after radiotherapy were used for analysis. Figure 11 shows the relation of jitter and shimmer with the overall impression of hoarseness as expressed by G.

Figure 11: Relation of jitter and shimmer with the overall impression of hoarseness



The relation of jitter and shimmer with the overall impression of hoarseness as expressed by G (mean of two raters) of the GRBAS scale. A total of 61measurements in 33 patients with head-and-neck cancer before and after radiotherapy are shown. Origin: Brockmann-Bauser and Drinnan, 2011.

As expected, jitter and shimmer are lowest in patients with G0 or G0.5 voices, where one or both raters scored G=0. Practically, these patients define the normal range for jitter and shimmer. Equally, some patients rated G1 and G2 have very high jitter or shimmer, and therefore the population mean for jitter (shimmer) is higher than for G0. Nevertheless, in our clinical work we are concerned with the individual. Yet Figure 11 clearly shows that the majority of pathologic voices have jitter and (especially) shimmer within the range of normal sounding voices.

Further, higher jitter or shimmer are not associated with an increased level of hoarseness, which one would expect from useful measures of perceptual dysphonia. In the present results it is not possible to distinguish between G1 and G2 voices on the basis of jitter or shimmer. This might help explain the contradictory evidence in the literature. In the present study we would have reported a moderate correlation ($r=0.25$) between shimmer and GRBAS G. But had we pre-selected our patients for evidence of dysphonia (i.e. inclusion criteria of G1 or higher), this correlation would disappear ($r=0.03$).

Based on the presented data and the contradictory evidence from the literature we have to conclude that jitter and shimmer do not accurately or reliably measure perceptual dysphonia per se. Therefore, both indices cannot be simply considered as objective measures for perceptual dysphonia. However, jitter and shimmer might give clinically useful measurements of subtle changes in mildly dysphonic or normal sounding voices (rated as G1 or below). But given that the relation with vocal fold pathology is unclear, it has to be established what jitter and shimmer might indicate in these patients, and whether it is clinically relevant.

2.3.4 *Are jitter and shimmer useful clinical outcome measures?*

As described in section 2.1 both jitter and shimmer have been frequently used as outcome measures following a variety of interventions to treat voice pathology (Carding et al., 2009). Given the criticism raised that they do not reliably indicate vocal fold pathology or perceptual dysphonia, one might suspect that jitter and shimmer are not useful outcome measures at all. This may not be true. In outcome measurements the effects of the naturally large differences between individual voices are minimised. This might help to identify useful applications of jitter and simmer. To investigate the effects of a treatment, measurements are usually made before and after the intervention. When an appropriate statistical test is used (typically Wilcoxon, paired t-test, or repeated measures ANOVA), each subject acts as their own control. Inter-subject differences become less important and the relation of jitter (or shimmer) with pathology only need be true *within the patient*, not across the population.

For use of jitter and shimmer as outcome measure, the evidence is more encouraging and broadly consistent. For example, in patients with unilateral vocal fold paralysis or sulcus vocalis, reduced instrumental acoustic perturbation indicated positive voice changes after surgical treatment in a number of studies (Lee et al., 2010; Zhang et al., 2010; Cheng et al., 2009; Hartl et al., 2009). Lee et al. assessed 124 patients receiving injection thyroplasty and reported significantly lower jitter and shimmer along with improved perceptual hoarseness and vocal fold closure (Lee et al., 2010). Hartl et al. reported a long term improvement in jitter and shimmer, two years after 94 patients received injection or Gore-Tex thyroplasty (Hartl et al., 2009).

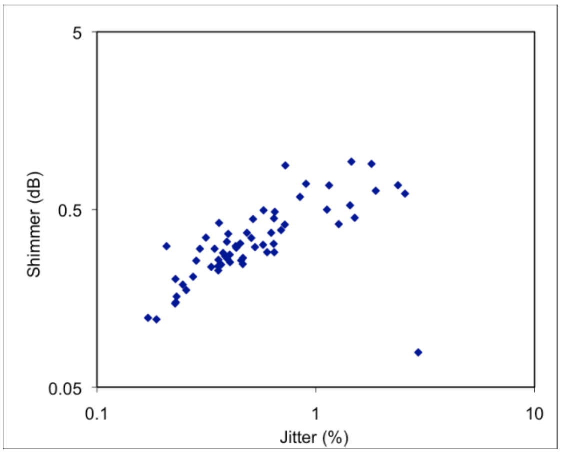
Based on this we conclude that jitter and shimmer are able to indicate differences *within* voices, but not *between* voices. However, it is not clearly established yet what jitter and shimmer indicate. This has to be clarified by further research for a valuable and routine clinical use as outcome measure.

2.3.5 ***Are jitter and shimmer independent from each other?***

Jitter and shimmer are normally considered as independent parameters contributing different information about vocal function (Dejonckere et al., 1996). Yet they are known to co-vary (Baken and Orlikoff, 2000c; Heilberger and Horii, 1982), and so the question arises: are jitter and shimmer essentially measuring the same phenomenon?

To the best of my knowledge this has not been investigated in voice patients to date. Therefore, the data presented in section 2.3.3 (Figure 11) will be further analysed with regard to the correlation between jitter and shimmer in dysphonic voices. Figure 12 shows the relation of jitter with shimmer (plotted on logarithmic scales) in 33 patients with head-and-neck cancer before and after radiotherapy. Clearly, there is a very strong relationship between jitter and shimmer (r=0.63). These preliminary results should be confirmed by a larger clinical study, since this would have considerable implications for the use of jitter and shimmer as independent measures of vocal impairment. However, based on this data we can conclude that jitter and shimmer clearly tend to covary in dysphonic voices. This needs to be considered when interpreting acoustic assessment results.

Figure 12: Relation of jitter and shimmer in 33 voice patients



The relation of jitter with shimmer in 33 patients with head-and-neck cancer before and after radiotherapy. The analysis is based on the same data as used in Figure 11 and was plotted on logarithmic scales. Origin: Brockmann-Bauser and Drinnan, 2011.

2.3.6 ***Summary***

Even though the evidence is clearly limited, acoustic analysis is widely used to assist differential diagnosis, documentation and evaluation of treatment for clinical voice disorders.

To date clinical studies have not proven that jitter and shimmer are absolute or independent indices of voice pathology or perceptual hoarseness. In comparisons within patients, jitter and shimmer might have value as an outcome measure. However, the validity of acoustic assessments in clinical applications has not been satisfactorily established.

2.4 Measurement reliability in voice clinics

2.4.1 *Why should we consider measurement reliability?*

In the preceding section, the validity of jitter and shimmer in current clinical voice diagnostics was questioned, but both measures may not be applied optimally. The key to establishing the validity of an assessment tool is reliability. A reliable method gives the same result irrespective of the examiner or variations in measurement conditions. Therefore, in the present section the reliability of jitter and shimmer measurements will be reviewed.

2.4.2 *Intraclass reliability: are jitter and shimmer measurements comparable between studies?*

Computer analysis packages often use similar labels such as “Jitter percent” and “Percent Jitter” suggesting comparability for different jitter and shimmer measures. However, a comparison of three analysis systems in 20 (Karnell et al., 1995) and 50 participants, respectively (Bielamowicz et al., 1996) showed a poor agreement between similar jitter and shimmer measures in vowel phonation despite comparable recording and analysis procedures. Also, a recent study by Maryn et. al. in 50 patients showed that the type of recording equipment and the computer analysis approach significantly influenced jitter and shimmer, even though current standards for acoustic assessments were applied (Maryn et al., 2009).

This low intraclass reliability was explained by a different system to acquire the voice samples and unequal software based waveform analysis and calculation strategies (Boersma, 2009; Maryn et al., 2009). Therefore, despite the use of similar labels, jitter and shimmer equality between different analysis systems cannot be assumed. Given this, pathology thresholds are also not transferable between analysis systems and programs (Boersma, 2009; Maryn et al., 2009). This hinders the comparison of assessments results between clinical centres, and also the meta analysis of data between studies.

2.4.3 *How reliable are jitter and shimmer in dysphonic voices?*

Previous work has shown that voice perturbation reliability is strongly influenced by voice signal overall regularity (Titze, 1995). Test-retest reliability for jitter and shimmer was at best moderate in dysphonic, but was better in 50 normal voices (Carding et al., 2004). In non-dysphonic voices only a moderate retest reliability within subjects was found (Dixon, 1999). This is mainly due to the fact that jitter and shimmer both depend on the correct recognition of F_0 and SPL (please see also sections 2.1 and 2.2), which is technically more difficult in irregular acoustic waves. Therefore, Titze subclassified disordered voices into three main types according to irregularity severity, and published suitable acoustic assessment recommendations for each group (Titze, 1995). These recommendations are summarised in Table 2.

Table 2: Suitability of irregular voice types for instrumental acoustic analysis

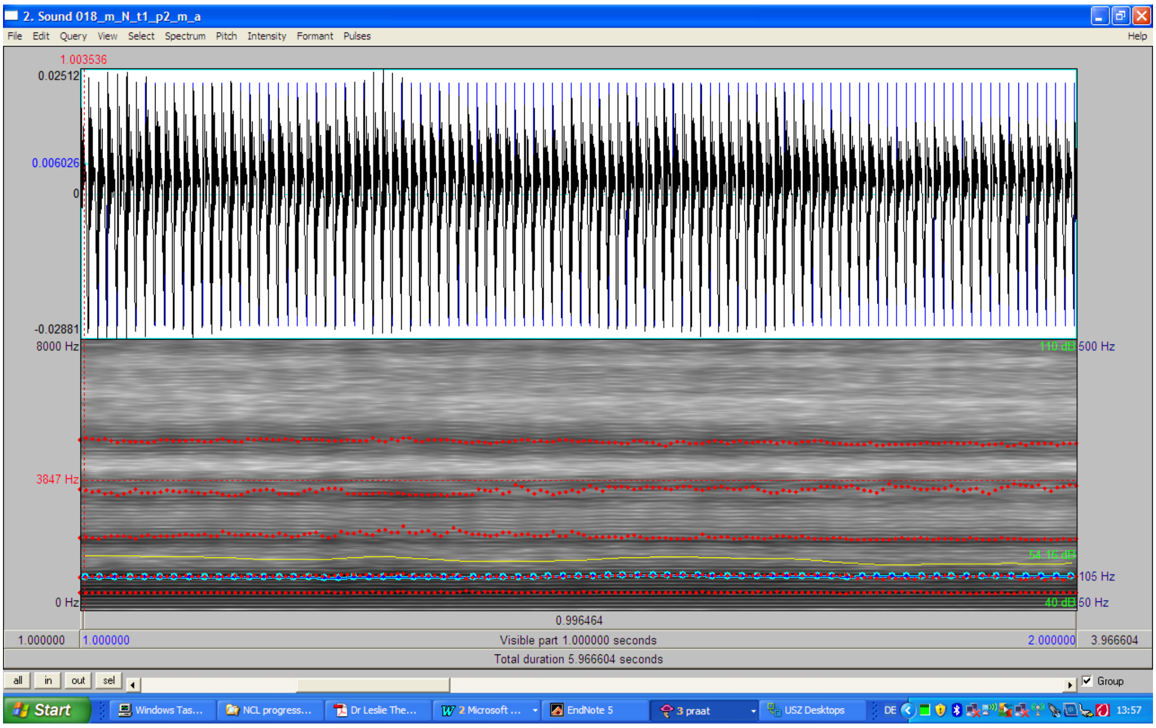
Voice classification	**Specifications according to Titze (1995)**	**Suitability for instrumental acoustic analysis**
Type 1 signal	Nearly periodic or normal voice signals	High suitability
Type 2 signal	Voices with qualitative changes or without obvious single F_0	Low suitability, spectrograms recommended
Type 3 signal	Chaotic or random voice signals	Unsuitable for instrumental acoustic assessments

Classification of voice types according to Titze (1995) and their respective suitability for instrumental acoustic analysis. Origin: own table.

Perturbation analysis was shown to be most useful and reliable in *Type 1* voices presumably tracking aperiodicities near or below listener's thresholds (Carding et al., 2004; Bielamowicz et al., 1996; Rabinov et al., 1995; Titze and Liang, 1993). In *Type 2* voice signals visual displays like spectrograms were suggested to examine pitch breaks and stability, the overall harmonic structure and subharmonics (Baken and Orlikoff, 2000c, Titze, 1995). In *Type 3* voices, jitter reliability has been shown to decrease considerably (Bielamowicz et al., 1996) whereas perceptual analysis reliability increases (Rabinov et al., 1995). Therefore perceptual voice ratings were recommended for the clinical assessment of Type 3 voices.

Below are examples of Type 1 (Figure 13) and Type 3 (Figure 14) voices, as displayed by the instrumental acoustic analysis software PRAAT (Boersma and Weenink, 2006). In the upper section of the window is an oscillogram, and in the lower part a spectrogram of the acoustic voice signal. The vertical blue lines in the upper window section and the horizontal dotted blue line in the lower section indicate the fundamental frequency (F_0) as determined by PRAAT. In Figure 13, showing a normal sounding voice, the recognition of F_0 is correct. By comparison Figure 14 displays the acoustic wave of a moderately dysphonic voice. In the perceptual assessment by GRBAS scale the voice has been classified as G2R2B1A1S2. In the more "rough" sounding section of the voice sample the software PRAAT is not able at all to determine F_0 (Figure 14).

Figure 13: Example of a healthy voice classified as Type 1 voice signal

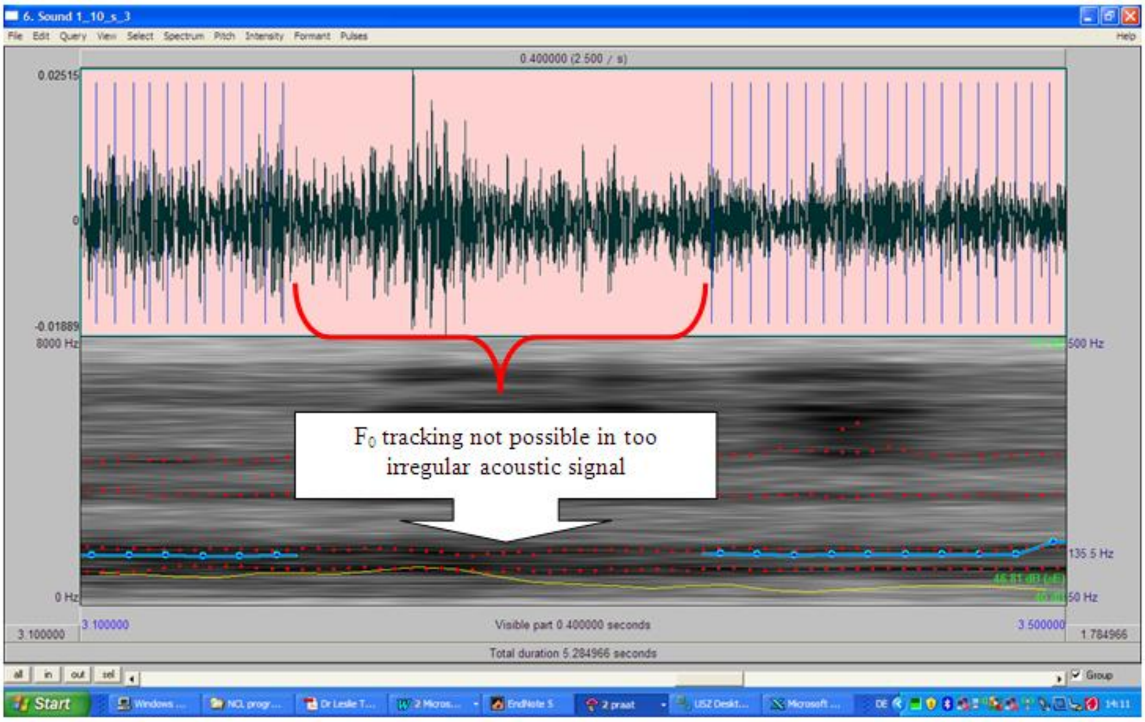


Screenshot of Type 1 voice signal as displayed by the software PRAAT. The upper window section shows the oscillogram. In the lower section is the sonagram with tracking of the fundamental frequency cycles, as indicated by blue dots. Origin: own graph.

Normally, dysphonic patients with more irregular voice signals are examined in voice clinics. Therefore, the described technical limits have enormous consequences for the practical applicability and usefulness of clinical perturbation measurements. In work investigating a sample of 181 unselected dysphonic outpatients approximately 20%

were not analysable at all by the analysis software Multi-Dimensional Voice Program (KayPentax, 1993) due to voice irregularity (Carding et al., 2004). This shows that even in voice clinics, a considerable proportion of the target group, dysphonic patients, cannot be assessed by instrumental acoustic analysis. Further, a study of 202 voice patients showed that only 42% of the disordered voices can be classified as Type 1 voice signals suitable for instrumental acoustic analysis at all (Behrmann et al., 1998). Thus, in the majority of voice patients acoustic analysis is not able to provide reliable measurements (Ma and Yiu, 2005).

Figure 14: Example of a moderately dysphonic voice classified as Type 3 signal



Type 3 voice signal processed by the software PRAAT. As indicated by the partially missing blue dots, the software is unable to determine the fundamental frequency in the more irregular sections of the acoustic signal. Origin: own figure.

This might explain why the literature regarding the validity of jitter and shimmer to indicate voice pathology or dysphonia is so inconclusive to date. Maryn et al. reported that the system to process voice recordings affects acoustic measurements by factors ranging from 1.2 to 3.1 (Maryn et al., 2009). To evaluate the clinical importance of these effects they have to be considered in the context of the changes we hope to detect in our patients. For example, in patients with polyps, shimmer was 1.6–1.8 times higher than in healthy voices (Jiang et al., 2009). Similarly, Lee et al. reported a reduction in jitter and shimmer by factors of 1.5–1.75 after thyroplasty (Lee et al., 2010). By comparison, the 1.2-fold to 3.1-fold effect due to the analysis system seems large, and we might expect it to have noticeable confounding effects. Indeed, as previously described, Jiang et. al. drew opposing conclusions from identical recordings analysed with two different instrumental analysis programs (MDVP and CSpeech) (Jiang et al., 2009).

Therefore, a main methodological problem in establishing the validity of jitter and shimmer is that jitter and shimmer are not useful measures in Type 2 and Type 3 voices, i.e. the majority of clinical voice patients (Carding et al., 2004; Behrmann et al., 1998; Titze, 1995).

2.4.4 ***Summary***

The recognition of frequency and amplitude patterns in an acoustic wave is technically more difficult in signals with inherent variability due to background noise (e.g. dysphonia). Jitter and shimmer exactness is determined by the correct recognition of F_0 and voice SPL. Therefore, a reliable measurement of jitter and shimmer has been described as limited to nearly periodic or normal voice (Type 1) signals. However, around 58% of all patients have Type 2 and Type 3 voice signals. This major methodological problem might have limited previous research in the validity of jitter and shimmer.

Comparability between acoustic analysis systems and software, and thus between clinical centers and studies, cannot be assumed. Therefore, in the interpretation of acoustic assessment data the measurement technique and analysis as well as the quality of the voice recordings have to be considered.

2.5 Are confounding factors associated with the clinical assessment procedure?

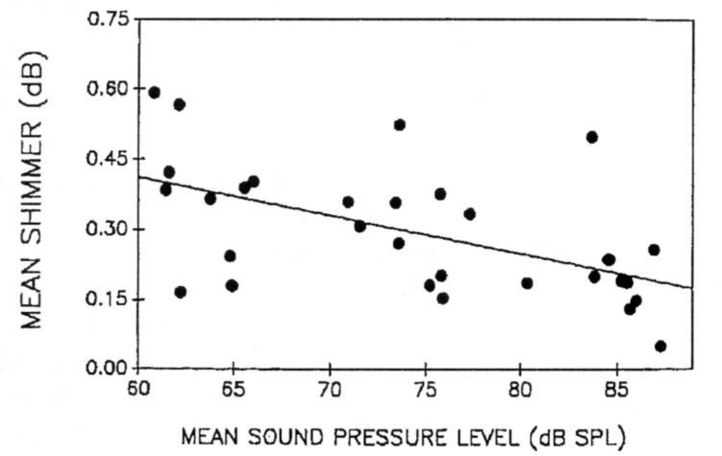
In clinical instrumental acoustic assessments, patients are usually instructed to phonate „at comfortable loudness and pitch“. However, patients do not always respond in the same way to this voice task. For example, Brown et. al. showed that women and men produce a significantly different voice SPL in this situation (Brown et al., 1996). Therefore, from a clinical perspective also the recording instructions and inter-individual differences in performing the given voice tasks might partially account for the high measurement variability. In the following sections potential confounding factors associated with the clinical measurement procedure will be reviewed in turn. According to Baken and Orlikoff these might include (Baken and Orlikoff, 2000c):

(a) the speaking voice intensity (voice SPL)
(b) the patient’s fundamental frequency (F_0)
(c) gender
(d) vowel choice

2.5.1 *Influence of voice Sound Pressure Level (voice SPL)*

Both jitter and shimmer have been shown to decrease with increasing voice SPL in a variety of works. For example Orlikoff and Kahane described a decrease of jitter and shimmer with increasing voice SPL in 10 healthy young men phonating /a/ at “stable pitch” in predefined voice SPL ranges between 60-68dB, 70-78dB and 80-88dB (Orlikoff and Kahane, 1991). Based on the data distribution a linear negative relationship between SPL and shimmer was assumed (Figure 15). Similar results were found for jitter and SPL.

Figure 15: Relation between voice SPL and shimmer

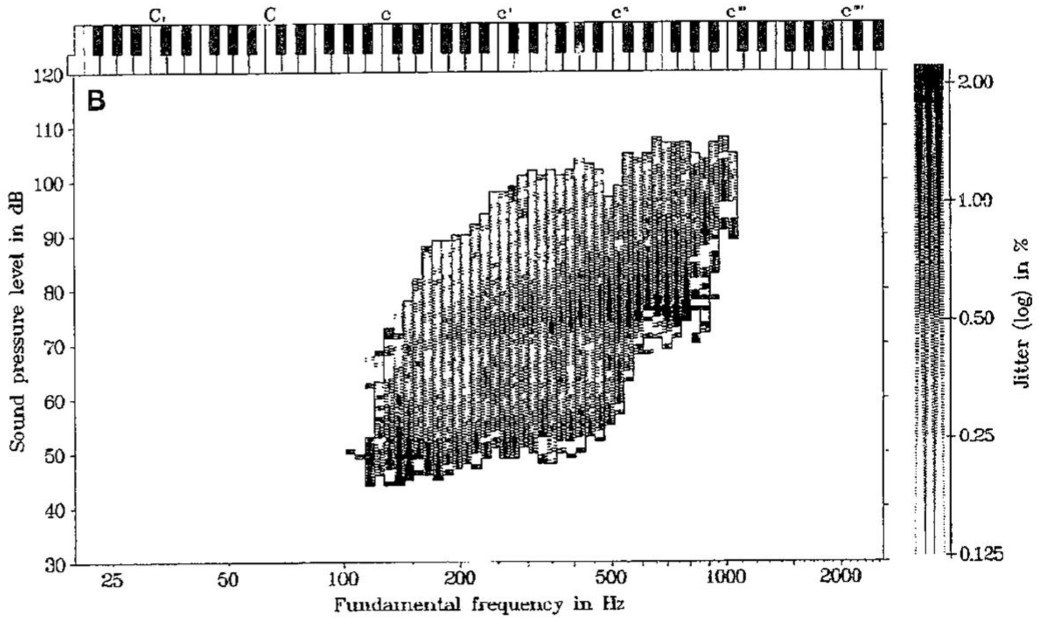


Relation between voice SPL and shimmer (dB) in 10 healthy men phonating at prescribed voice loudness levels. Origin: Orlikoff and Kahane, 1991.

This was also reported in 70 adults with muscle tension related dysphonia or superficial vocal fold pathology phonating at “comfortable loudness and pitch” as compared to “louder voice” (Dejonckere, 1998). Further, a general decrease of jitter and shimmer with increasing voice loudness was found in phonetograms, which were supplemented with perturbation measurements (Figure 16, Pabon, 1991; Pabon and Plomp, 1988) or in vowel phonations at prescribed voice SPL and F_0 levels (Gelfer, 1995). Figure 16 illustrates the relation of jitter (as indicated by the colour intensity) with voice SPL (y-axis) and F_0 (x-axis) averaged for 9 women in a phonetogram (Pabon, 1991).

Based on these studies it can be concluded that both jitter and shimmer are affected by changes in voice SPL. In voice clinics this might be a substantial confounding factor, since habitual voice SPL varies considerably between individuals, but also between gender or age groups (Brown and Shrivatsav, 2007; Hodge et al., 2001; Brown et al., 1996).

Figure 16: Phonetogram supplemented with jitter measurements



Averaged phonetogram (= "voice range profile") for 9 women. The y- axis shows voice intensity (dB) and the x-axis fundamental frequency (Hz). The extent of jitter (%) in vowel phonation is indicated by the colour intensity. Origin: Pabon, 1991.

However, a transfer of these conclusions to a successful control of voice SPL effects in voice clinics is not possible due to a number of reasons: (a) in some studies the applied jitter and shimmer parameters have not been specified, (b) jitter and shimmer have been averaged for large or undocumented SPL ranges (Dejonckere, 1998), (c) the examined groups have been small and highly inconsistent (Dejonckere, 1998; Orlikoff and Kahane, 1991; Pabon, 1991; Pabon and Plomp, 1988), (d) to date it is unclear if it affects jitter or shimmer, when patients are asked to control for their voice SPL and/or F_0 (Gelfer, 1995; Orlikoff and Kahane, 1991).

2.5.2 ***Fundamental frequency influence***

The literature base describing the effects of fundamental frequency (F_0) is sparse and shows contradictory results (Baken and Orlikoff, 2000a; Baken and Orlikoff, 2000b). From section 2.2.3 we are aware that absolute jitter values are smaller in voices with higher F_0. However, comparisons between healthy women and men with a *relative* jitter measure showed an association between a higher F_0 in women and increased jitter (Fitch, 1990). In turn an investigation in women and men changing their fundamental frequency (F_0) found that relative jitter and shimmer were highest in lower frequencies (Gelfer, 1995). This was also the case in a study applying an absolute jitter measure, where absolute jitter decreased more in the men than in women (Orlikoff and Baken, 1990).

Since few studies have described fundamental frequency effects, Baken and Orlikoff concluded that the influence of F_0 on jitter and shimmer is not fully understood to date (Baken and Orlikoff, 2000a; Baken and Orlikoff, 2000b). However, based on the current evidence we have to expect that an influence of F_0 might underlie gender and vowel effects. Further, in vocally healthy adults F_0 has been shown to normally increase with rising voice SPL (Gramming, 1988). Therefore, the combined effects of F_0 with gender, vowel or voice SPL might also influence jitter and shimmer. Thus, changes in F_0 might be a considerable confounding factor in clinical acoustic assessments.

2.5.3 ***Gender differences***

In the research literature, gender effects have been described in highly contradictory terms. Comparing two similar designed studies in 31 men and 20 women (Sorensen and Horii, 1983; Horii, 1980), Sorensen and Horii concluded that female voices normally display less shimmer but more jitter than male voices (Sorensen and Horii, 1983). This hypothesis was supported by later studies (Fitch, 1990; Deem et al., 1989). Also, when comparing /a/, /i/ and /u/ in 25 women and 24 men, significantly higher jitter was reported for women across all vowels (Dwire and McCauley, 1995).

However, also smaller “Absolute Jitter” values were found in women as compared to men (Sussmann and Sapienza, 1994; Jafari et al., 1993). In contrast to this, similar jitter was reported in 6 women and 6 men phonating at controlled voice SPL and F_0 (Orlikoff and Baken, 1990). Also, in this study men had a bigger change in jitter than women when they increased their fundamental frequency (F_0).

The reported studies are difficult to compare to each other because of (a) small numbers of participants, (b) phonations at unequal voice SPL and F_0 levels, (c) the use of different voice analysis tools. Further, the reported study results from Orlikoff and Baken suggest that gender effects might be linked to the individuals' voice SPL and F_0 (Orlikoff and Baken, 1990). Therefore, we have to conclude that physiologic gender differences in measurements of jitter and shimmer have not been well understood to date. This hinders an efficient control of gender effects in voice clinics.

2.5.4 ***Vowel effects***

Vowel effects in measurements of jitter and shimmer have been investigated in a number of studies. Table 3 gives an overview of the main findings in vocally healthy adults and the methodologies used. Again, the reported results are highly contradictory and range from the description of significant vowel differences (Wilcox and Horii, 1980) to no distinct vowel effects (Orlikoff, 1995). Highest jitter has been found in /u/, /i/ or /a/, and lowest shimmer in /i/ or /u/ (Table 3). These clearly contradictory results show that the influence of vowel articulation on jitter and shimmer is not fully understood to date.

From a methodological standpoint the reported results appear not comparable or transferable into clinical practice due to a number of reasons such as (a) unequal recording equipment between studies not complying to current guidelines, (b) the use of both male and female subjects in the same analysis (Orlikoff, 1995; Milenkovic, 1987), (c) measurements at different voice intensities and fundamental frequencies (Delyiski et al., 2006; Titze, 1995), (d) an unequal recording distance (summarised in Table 3). However, as discussed in the previous sections, all of these factors have to be considered when determining the relative influence of different vowels on jitter and shimmer. Therefore, we have to conclude that the effects of vowels in clinical measurements of jitter and shimmer have not been satisfactorily described to date.

Table 3: Literature background regarding vowel effects on jitter and shimmer

Reference	Highest jitter	Lowest jitter	Highest shimmer	Lowest shimmer	Participants	Recording conditions
Kilic 2004	/u/ and /y/		/a/, /e/, /u/, /y/		26 men	"comfortable loudness and pitch"; recording distance 15cm
Dwire & Mc Cauley 1995	/u/ * (only for women)				25 women/ 24 men	"comfortable loudness and pitch"; patient held microphone; recording distance 2,5- 3,5cm
Gelfer, 1995	/i/ and /a/ equal		/a/*	/i/*	29 women	Speaking F_0 and speaking F_0 + 1 octave; 60, 70 and 80dB; distance 12 inches
Sussm. & Sapienza 1994	/a/* and /i/*	/u/*			10 women/ 10 men	"comfortable loudness and pitch"; recording distance 15cm
Sorensen& Horii 1983	/i/	/a/	/a/	/u/	20 women	"comfortable F_0"; between 70 and 80dB; recording distance 15cm
Wilcox& Horii 1980	/a/* and /i/*	/u/*			20 young/ 20 aged men	"comfortable F_0"; between 70 and 80dB; recording distance 15cm
Orlikoff 1995	/i/ and /u/	/a/	/i/	/a/	10 women/ 10 men	Women 220Hz & men 110Hz +/- 0.5 semitones; 74 +/- 4dB; recording distance 30cm

Research in vowel differences in healthy speakers summarized according to highest and lowest jitter and shimmer values. The reported study results are clearly contradictory and range from the description of significant to low vowel effects. This might be partly due to different recording conditions. The effects marked with an * were reported as statistically significant. Origin: Brockmann et al., 2011.

From a theoretical perspective, various hypotheses have been proposed to explain or reject the possibility of vowel effects on voice intensity and fundamental frequency perturbation. According to the source-filter theory (section 1.1.5), jitter and shimmer should not be affected by vowel articulation (Fant, 1980). Alterations in the vocal tract shape have been described to influence the overtone spectrum (filter function), but not to alter the acoustic signal of the sound source (larynx) (Kent, 1993). This hypothesis is supported by electroglottographic (EGG) measurements of jitter and shimmer showing no distinct vowel effects (Orlikoff, 1995).

However, the *physical linkage hypothesis* proposes that movements in the muscles of the vocal tract and the larynx might be related to each other. Indeed, investigations in the position of the hyoid-larynx complex (Honda, 1983) or the head and tongue (Lin et al., 2000) showed that vowel articulation was associated with changes in fundamental frequency and also the amount of jitter and shimmer. It may be argued that changes in jitter and shimmer might be a simple side effect of vowel specific alterations in F_0 and probably also voice SPL. However, a change in supraglottal muscle activity during articulation has been shown to be associated with alterations in laryngeal muscle activity (Lin et al., 2000). Modifications in laryngeal muscle tone could influence the vibratory properties of the vocal folds and hence also jitter and shimmer (Brockmann et al., 2008). For example, when the tongue is in an elevated position, the activity in the geniohyoid and laryngeal strap muscles is higher (Lin et al., 2000). This pulls the larynx upward and forward and causes a tilt of the thyroid cartilage. By this the vocal folds are indirectly stretched, resulting in a higher vocal fold tone, increased F_0 and probably even reduced jitter and/or shimmer (Hirano, 1981; Lin et al., 2000).

2.5.5 ***Summary***

At present the effects of voice SPL, gender, vowel and F_0 have not been satisfactorily investigated and described. Based on the current literature it is not possible to efficiently control for each of these factors in clinical measurements of jitter and shimmer.

In clinical assessments, patients are commonly asked to say a prolonged vowel at “comfortable loudness and pitch”, though women and men phonate at an unequal voice SPL in this situation. Therefore, in clinical measurements gender effects might be distorted by the influence of systematically different voice SPL between women and men. Similar combined effects might also be present for voice SPL and F_0, F_0 and gender, or even F_0 and vowel. Acoustic analysis programs do not provide SPL, F_0 or vowel specific normative thresholds. Therefore, all single factors or their combined effects might partially account for the limited measurement sensitivity, reliability and specificity in voice clinics.

2.6 Why should we improve clinical jitter and shimmer measurements?

If the effects of SPL, F_0, gender and vowel are significant, failure to consider them in clinical voice assessment provides a false picture of the acoustic properties of the patient's voice. This might delay the patient's access to appropriate further diagnostics and treatment and, in the worst case, even support false diagnoses.

In turn, acoustic analysis could have unexplored potential in tracking voice changes at a very fine level below listener's thresholds. For example, in voices that are not classified as pathologic in perceptual assessments, jitter and shimmer might provide more in-depth information about aberrant acoustic patterns. Also, the subjective phenomenon of "normal voice" might be described in more detail by jitter and shimmer. In clinical practice this would be highly valuable in a number of applications:

(a) Discrete alterations in essentially normal sounding voices might be associated with muscle tension dysphonia, which is difficult to diagnose by other examination techniques (Altman et al., 2005; Schneider et al., 2002; Dejonckere et al., 2001).

(b) Subtle voice changes might indicate early neoplasm growth, in which cases an early diagnosis is associated with a significantly better treatment outcome and quality of life (Licitra et al., 2003).

(c) Jitter and shimmer might help to assess intervention success at an early stage, which would make interventions more efficient.

(d) Acoustic assessments are non-invasive and provide information independent of the examiner about discretely aberrant acoustic voice patterns. Therefore, instrumental analysis might be a useful clinical screening tool.

2.7 **Conclusions and main aims of present PhD work**

In the present chapter it has been argued that clinical jitter and shimmer measurements are far from being reliable and objective voice assessment tools. This is due to a number of confounding factors associated with the measurement technique and the acoustic assessment protocol.

An analysis of the current literature shows that the effects of voice SPL, F_0, gender and vowel have not been sufficiently described for an efficient control in clinical assessments. Furthermore, our understanding of what jitter and shimmer might indicate is too limited for a pathology specific or valid interpretation of acoustic assessment results. In turn, jitter and shimmer might have unexplored diagnostic potential in subtle or early voice pathology (such as muscle tension dysphonia). With this background in mind the main study questions of the present work are defined:

(1) How important are the effects of voice SPL, F_0, gender and vowel in clinical measurements of jitter and shimmer?

(2) How can we improve clinical measurements of jitter and shimmer?

(3) What can clinical jitter and shimmer measurements tell us about the human voice?

(4) What are useful applications for jitter and shimmer measurements?

Chapter 3. Main methodology

In the present thesis, several experimental studies are described in chapters 4 – 8. These were designed to address the main study questions defined in the preceding chapter 2. The present chapter 3 summarises selection criteria and the recruitment approach for all study participants examined. Also specified are the main recording and voice signal analysis techniques, as well as the measures to ensure data quality. Chapters 4 - 8 refer back to the present chapter in their methods section.

3.1 Main exclusion criteria for study participants

3.1.1 *Experiment chapters 4, 5, 6 and 7*

For the experiments described in chapters 4, 5, 6 and 7 the same participant exclusion criteria were used to maintain data comparability. All participants were vocally healthy volunteers recruited via e-mail from the University Hospital Zurich. The exclusion criteria were assessed by a questionnaire developed to cover the main known factors influencing vocal health or jitter and shimmer (Appendix G). Information about the subjects' smoking habits and native language were collected, but not used for further analysis.

Exclusion criteria:
(a) a hoarse voice on the day of recording as assessed by GRBAS scale (GRBAS<1);
(b) recent voice problems or a voice disorder history;
(c) any previous formal voice training or voice therapy;
(d) a medication or a condition that might affect normal voice function;
(e) recent intubation for any surgical intervention;
(f) surgery in the torso, head and neck region in the last 18 months.

Furthermore, three recorded phonations of /a/ were perceptually analysed by two independent voice experts (one speech-language pathologist and one phoniatrician) using the GRBAS scale (Hirano, 1981). Please refer to section 3.2 for a detailed description of inclusion criteria for voice experts.

Voice recordings were excluded from objective acoustic analysis if the mean of the two expert ratings was 1 or higher for any GRBAS characteristic. This was done to ensure that the voice samples were all Type 1 voice signals and thus suitable for instrumental acoustic analysis (please see also section 2.4).

3.1.2 *Experiment chapter 8*

In the study described in chapter 8, teachers were recorded at their workplace in schools. Since very similar exclusion criteria were applied as described above, the same questionnaire was used to assess the volunteers. The differences to the preceding studies were:

- Teachers with previous voice training experience were admitted to the study since the majority of teachers in Switzerland have voice training during their formal education.
- Teachers with previous voice therapy or with episodes of hoarseness were admitted to the study. This was allowed since the prevalence of perceived "hoarseness" or voice disorders is extremely high in teachers as compared to other professions (Bermúdez de Alvear et al., 2010; Roy et al., 2004).
- Only teachers with an overall GRBAS grade <1 as determined by the examiner were admitted to the study. In this study, all voices passing this criterion were included for instrumental acoustic analysis.

3.2 **Inclusion criteria for voice experts**

In the studies described in chapters 4, 5, 6, 7 and 8 "voice experts" were involved in perceptual analysis tasks. All were staff from the Department of Phoniatrics and Speech Pathology, University Hospital Zurich, Switzerland.
Inclusion criteria:

- Professional training as speech pathologist or phoniatrician
- Work experience (diagnostics and therapy) with voice disordered patients of 2 years minimum
- Participation in regular GRBAS training session (2 per year) at the Department of Phoniatrics and Speech Pathology, University Hospital Zurich

3.3 **Recording methods**

As discussed in section 2.2, the recording method, equipment and environment as well as the computer analysis technique significantly influence measurements of jitter and shimmer. Therefore, the measurement method and equipment complied as close as possible to current technical and clinical guidelines (section 2.2.6; Dejonckere et al., 2001; Titze, 1995). The present chapter describes the general methods used in the experiments of the present thesis. To ensure data comparability, a similar recording and analysis technique was applied for all voice recordings.

3.3.1 ***Microphone choice***

In all experiments, the same head-mounted microphone was used (AKG Acoustics, C444). It was always placed in an off-axis position with 10cm microphone-mouth distance (Titze, 1995). To minimise the effects of background or breathing noise a protective styrofoam cap was used (Figure 17).

Figure 17: Recording setting in a sound proof room

Recording setup in a sound proof room at the University Hospital Zurich. Origin: own photograph.

3.3.2 *Equipment chapters 4, 5 and 7*

Figure 17 shows the recording equipment and set up as used for the experiments described in chapters 5, 6 and 8. All participants were recorded with a portable DAT-Recorder (Sony, TCD-D8) in a sound-proof room with an ambient noise of approximately 20dBA. The recordings were later converted to digital sound files (.wav format) with a sampling rate of 48000Hz and 16-bit resolution. Later, each individual phonation was cut out and anonymously labelled on a laptop with the software Audacity 1.2.4b (Mazzoni et al., 2005).

Recording equipment list:

(a) microphone (AKG Acoustics, C444); and
(b) a portable DAT recorder (Sony, TCD-D8).

3.3.3 *Equipment chapter 6*

In the research experiment described in chapter 6 healthy adults were recorded, while they performed a range of voice tasks. The technical voice recording equipment is identical to studies 4, 5 and 7. Further, an additional monitor was attached to the computer. When the participants phonated, the sound pressure level of their voice was shown on the extra monitor. Coloured arrows marked the target intensity levels of 65, 75, 85 and 95dBA (please see chapter 6 for details). Thus, the study participants were able to control for their voice intensity by visual feedback. Also, in this study a piano was used to provide auditory feedback.

Equipment list:

(a) microphone (AKG Acoustics, C444); and
(b) a personal computer (Laptop); and
(c) an external soundcard (EMU 0404 USB 2.0 Audio); and
(d) the software Audacity version 1.3.6 beta, set at a sampling rate of 48000Hz and 16-bit quantization (Mazzoni et al., 2008); and
(e) an additional monitor marked with coloured lines indicating the target voice SPLs of 65, 75, 85 and 95dBA; and
(f) a piano.

3.3.4 *Equipment chapter 8*

In this study, vocally healthy teachers were recorded during a working day at the school. Further, they were asked to control for voice SPL.
Thus, all recordings were done in normal room acoustics using the following equipment:

(a) microphone (AKG Acoustics, C444); and
(b) a personal computer (Laptop); and
(c) an external soundcard (EMU 0404 USB 2.0 Audio); and
(d) the software Audacity version 1.3.6 beta, set at a sampling rate of 48000Hz and 16-bit quantization (Mazzoni et al., 2008); and
(e) an additional monitor marked with a red line indicating the target voice SPL of 85dBA.

3.4 **Calibration and voice analysis technique**

3.4.1 *Calibration of the recording system and calculation of voice SPL (dBA)*

Conversion from *uncalibrated voice signal amplitude* as measured by PRAAT to *calibrated SPL values* was done by comparison method (Winholtz and Titze, 1997a). The identical procedure was used in all experiments.

Prior to voice recordings, calibrated speech weighted noise (Wagener et al., 1999) was recorded with 10cm distance to the sound source. Depending on the experiment the Sound Pressure Levels (dBA) ranged from 50dBA to 95dBA. Thereafter, the difference between known SPL (dBA) from the calibrated signal and the measured uncalibrated amplitude values was calculated. The difference was later used to compute calibrated voice SPL (dBA) of the voice recordings.

3.4.2 *Instrumental acoustic analysis: settings in the software PRAAT*

Instrumental acoustic analysis was done with the specialised free instrumental analysis software PRAAT. In chapters 4, 5, 6 and 7, version 4.4.16 was used (Boersma and Weenink, 2006). For the study described in chapter 8, version 5.1.03 was applied

(Boersma and Weenink, 2009). There was no change in the acoustic analysis strategies between both software versions, so data comparability can be assumed.

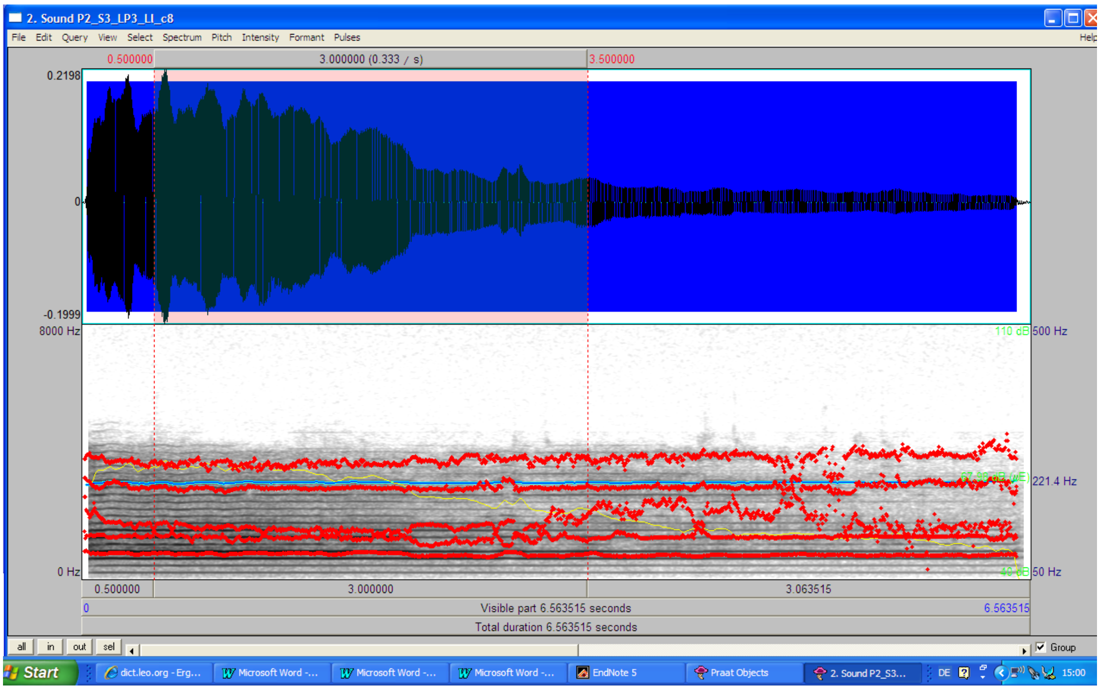
As discussed in detail in chapter 2.2, the strategy to extract the fundamental frequency (F_0) and the voice Sound Pressure Level (voice SPL) are key to successfully determining jitter and shimmer (Boersma, 2009; Maryn et al., 2009). PRAAT employs the waveform-matching method, which is considered to provide the most reliable results (Boersma, 2009; Maryn et al., 2009). Also, PRAAT offers a wide range of analysis strategies and settings. Table 4 summarises the main settings applied in all described experiments, and Figure 18 shows a representative screenshot of how the software was used.

Table 4: Settings for the menu items "Pitch", "Intensity" and "Pulse" in PRAAT

Menu	**Software settings**	**Reason (Reference)**
Pitch	Range: default (75-500)	Pitch of the human voice ranges from around 89 to 500Hz (Baken and Orlikoff, 2000c)
	Unit: Hz	Standard reporting unit (Baken and Orlikoff, 2000c)
	Analysis method: Autocorrelation	Autocorrelation method best to determine jitter (Boersma, 2009)
	Drawing method: default	The default settings have been recommended as most suitable for healthy voices (Boersma and Weenink, 2009)
Intensity	Range 40-110dB	The human voice is able to produce between 45 and 110dB (Baken and Orlikoff, 2000c)
	Averaging method: Mean energy, subtract mean pressure	PRAAT dB values are not calibrated, therefore the default setting was used to later determine calibrated voice SPL by comparison method
Pulses	Advanced pulse settings: default	The default settings have been recommended as most suitable for healthy voices (Boersma and Weenink, 2009)

Summary of main settings in the instrumental acoustic analysis software PRAAT. Origin: own table.

Figure 18: Screenshot of the software PRAAT as used for voice analysis



Representative image (screenshot) of how the instrumental acoustic voice analysis software PRAAT was used. Origin: own image.

3.4.3 *Choice of jitter and shimmer parameters*

In section 2.4.2 it was discussed that it cannot be automatically assumed that different jitter and shimmer types are comparable to each other. Therefore, in this section the details of the parameters chosen will be described. In all studies in this thesis, *F_0- and SPL- adjusted jitter and shimmer indices* were applied. Please refer to section 2.3.3 for a detailed discussion of the main different jitter and shimmer types.

In the studies presented in chapters 4 and 5, the parameters *jitter %* and *shimmer dB* of the software PRAAT were used (Boersma and Weenink, 2006). These specific parameters were chosen to allow the best possible comparison to key works about the effects of voice SPL and F_0 on jitter and shimmer from 1991 (Orlikoff and Kahane, 1991, Pabon 1991).

However, in more recent publications the indices jitter and shimmer are mainly calculated as a percentage (Baken and Orlikoff, 2000a; Baken and Orlikoff, 2000b). To ensure data comparability with present research, in chapters 6, 7 and 8 the indices *jitter %* and *shimmer %* were applied (Boersma and Weenink, 2006).

Shimmer dB is defined as the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20. *Shimmer %* is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude (Boersma and Weenink, 2006). The main difference is that shimmer dB is the ratio of the amplitudes of consecutive waves (expressed on a logarithmic scale), whereas shimmer % is the amplitude deviation, divided by the overall amplitude (Boersma and Weenink, 2006).

Table 5: Summary of jitter and shimmer measures per chapter

Chapters	**Jitter measure**	**Shimmer measure**
4 and 5	jitter %	shimmer dB
6, 7 and 8	jitter %	shimmer %

Summary of the applied jitter and shimmer parameters per chapter. Origin: own table.

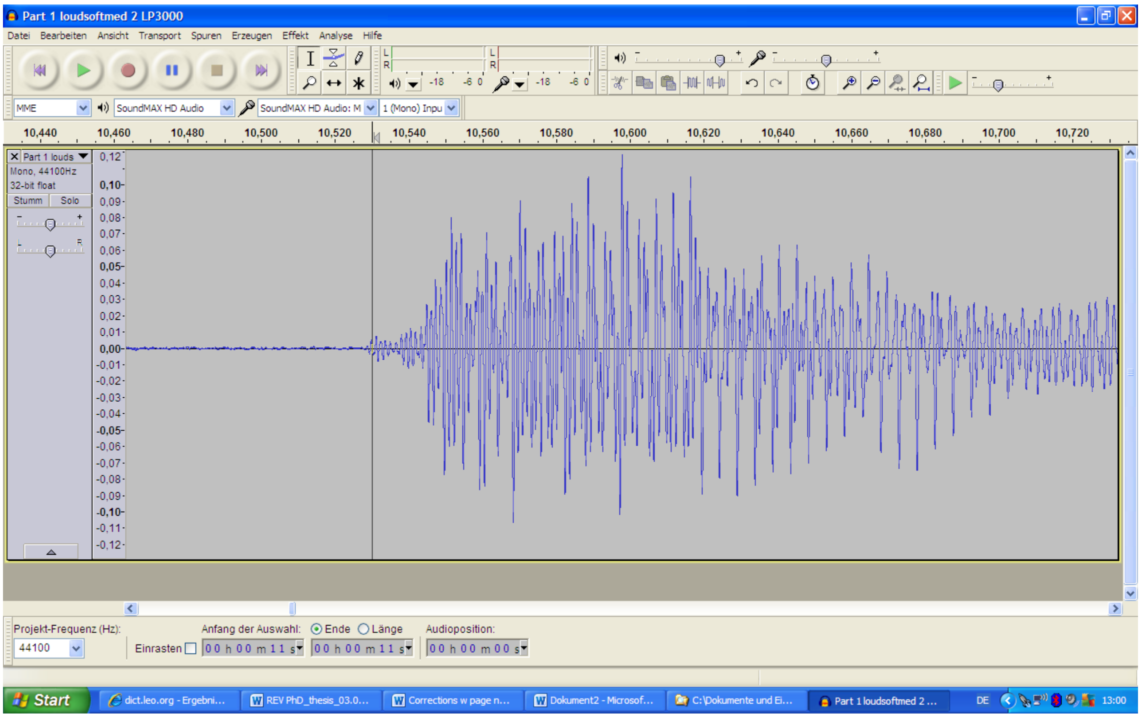
3.4.4 ***Preparation of voice recordings***

In all studies described in chapters 4, 5, 6, 7 and 8 instrumental analysis was conducted in vowel phonations. For this, each individual vowel phonation was displayed, cut out and anonymously labelled on a laptop with the software Audacity (please see section 3.3). Further, Audacity was used to filter all voice signals with a noise notch filter. For every voice sample, the filter was set at 50 Hz and Q 5 (Mazzoni et al., 2005).

3.4.5 ***Choice of acoustic analysis window***

Since the acoustic wave has been shown to exhibit greater random variability during the voice onset and offset phase, it has been recommended to use a stable portion of the voice signal for instrumental analysis (Dejonckere et al., 2003; Titze, 1995). In the studies of the present thesis a specific time frame excluding the onset and offset phase was always chosen. The timeframe was determined from the voice onset. In each voice recording the voice onset was defined as the first full acoustic wave in the oscillogram. For this, the voice signal was displayed in the software Audacity (Mazzoni et al., 2005). Figure 19 shows a representative image of the procedure. The definition of the voice onset was more difficult in recordings with very soft phonations and additional background noise, such as in the study described in chapter 8. Since the software Audacity allows a strong enlargement of the oscillogram none of the recordings had to be excluded due to difficulties in determining the voice onset.

Figure 19: Definition of the voice onset



Representative image (screenshot) of how the voice onset was determined using Audacity. Origin: own image.

In the studies described in chapters 4, 5, 7 and 8, second 0.5 to 3.5 after the voice onset were analysed. Thus the increased variability of the voice onset and offset phase were excluded from acoustic analysis.

However, in the experiment described in chapters 6, second 2.5 to 4.0 was chosen as analysis window, since the voice signals of some participants were not stable in fundamental frequency and/or voice SPL until second 1.3 from voice onset. This was the case when the participants had to control for voice SPL and pitch at the same time. Please refer to chapter 6 for further details on the study.

3.5 Quality control of instrumental measurements

3.5.1 *Quality control procedure*

A PRAAT script was used for instrumental acoustic analysis of the study data described in chapters 4, 5, 6 and 8. With the script, PRAAT automatically opened the sound files in a specific folder and analysed defined acoustic voice properties such as voice SPL, F_0, jitter and shimmer. Thereafter, the results were automatically written to a .txt file.

To ensure data correctness the analysis of the parameters voice SPL, F_0, jitter and shimmer was repeated "by hand" for a selection of sound files. From each data set the following recordings were chosen for a second analysis:

- 30 randomly chosen voice samples
- the 10 samples with highest voice SPL
- the 10 samples with highest F_0
- the 10 samples with lowest voice SPL
- the 10 samples with lowest F_0

All sound files were opened with the software PRAAT, displayed and visually inspected. The samples were checked for errors in the recognition of F_0 and voice SPL. Thereafter, the analysis window was selected and the instrumental analysis of voice SPL, F_0, jitter and shimmer was conducted. For this, the "Voice report" was chosen from the "Pulses" menu (Figure 18). These analysis results from the "Voice report" were then compared to the results of the original automated tests generated by the PRAAT script.

3.5.2 *Results of the repeated measurements*

In the studies reported in chapters 4, 5, 6 and 8, no relevant discrepancy between the automated and the repeated measurements was found. F_0 and voice SPL were generally correctly determined in the samples from the studies described in chapters 4, 5 and 6.

However, in the samples of the teacher study described in chapter 8, F_0 was not always correctly determined. This was the case for two recordings of one teacher with unusually soft phonations at “normal” voice intensity. In this study, all voice recordings were done at local schools in normal room acoustics. Therefore, the samples contained more random background noise, which lead to errors in the recognition of F_0. Please refer to section 2.2 for an in-depth discussion of the underlying technical reasons. In the data presented, jitter and shimmer were only affected to a minimal extent; the discrepancies were generally <0.01%. Therefore, the respective voice recordings were included into the analysis. These were perceptually analysed, and the correct F_0 was determined with the help of an electrical guitar tuner.

Chapter 4. Voice loudness and gender effects on jitter and shimmer

The work presented in chapter 5 has been published as follows (Appendix B):

Brockmann, M., Storck, C., Carding, P. N. and Drinnan, M. J. (2008) 'Voice Loudness and Gender Effects on Jitter and Shimmer in Healthy Adults', Journal of Speech, Language and Hearing Research, 51, (5), pp. 1152-60.

The literature review of chapter 2 showed that the influence of voice SPL and gender in clinical jitter and shimmer measurements has not been sufficiently understood for an efficient control. However, both factors might have a considerable impact on jitter and shimmer. In the present chapter, the confounding effects of voice SPL and gender will be assessed in a typical clinical voice task.

Chapter 4 summarises the MSc study from M. Brockmann "Voice Loudness and gender influence on Jitter and Shimmer in healthy adults" submitted to Newcastle University in August 2006. For further details, please refer to the paper based on this work (Appendix B).

4.1 **Introduction**

As discussed in section 2.5, both voice SPL and gender might be substantial confounding factors in clinical measurements of jitter and shimmer. In previous works, gender effects on jitter and shimmer have been described in highly contradictory terms (section 2.5). Women have been shown to display less shimmer and more jitter (Deem et al., 1989; Sorensen and Horii, 1983), but also smaller absolute jitter values than men (Jafari et al., 1993; Ludlow et al., 1987). Further, no significant gender differences were found (Orlikoff and Baken, 1990). Based on the above described results an efficient control of gender effects in voice clinics is not possible.

In contrast to this, similar voice SPL effects have been reported by several studies for healthy and pathologic voices: generally higher jitter and shimmer were found at lower voice SPL levels (Dejonckere, 1998; Orlikoff and Kahane, 1991; Pabon, 1991; Pabon and Plomp, 1988). However, as discussed in detail in section 2.5, these reports are limited for a useful clinical application due to a number of reasons:

(a) Jitter and shimmer have been averaged for large or undocumented voice SPL or F_0 ranges, or were measured at prescribed F_0 and voice SPL levels
(b) The acoustic parameters or statistical tests have not been specified
(c) The examined groups have been small and highly inconsistent

These issues were addressed in the design of the present study. For further details please refer to Appendix B.

4.2 **Main study aims**

The main aims of this study were to investigate the effects of voice loudness and gender on jitter and shimmer in a typical clinical voice task.

4.3 **Methods**

In a cross-sectional single cohort study 57 vocally healthy volunteers (28 women, 29 men) phonating the vowels /a/, /o/ and /i/ at individually “soft”, “normal” and “loud” voice were recorded in randomised order. Each intensity level (soft/normal/loud) was repeated three times, giving a total of 1539 single phonations for acoustic analysis.

For the master thesis summarised in this chapter, only the phonations with the vowel /a/, i.e. 513 single phonations, were analysed. The effects of phonation level (soft/medium/loud) and gender (f/m) on measured voice SPL (dB), F_0 (Hz), jitter (%) and shimmer (dB) were assessed with descriptive and inferential (ANOVA) statistics.

A publication based on this work can be found in Appendix B (Brockmann et al., 2008). Later, the data were further analysed with regard to the relative influence of the confounding factors voice SPL, F_0, gender and vowel (chapter 5).

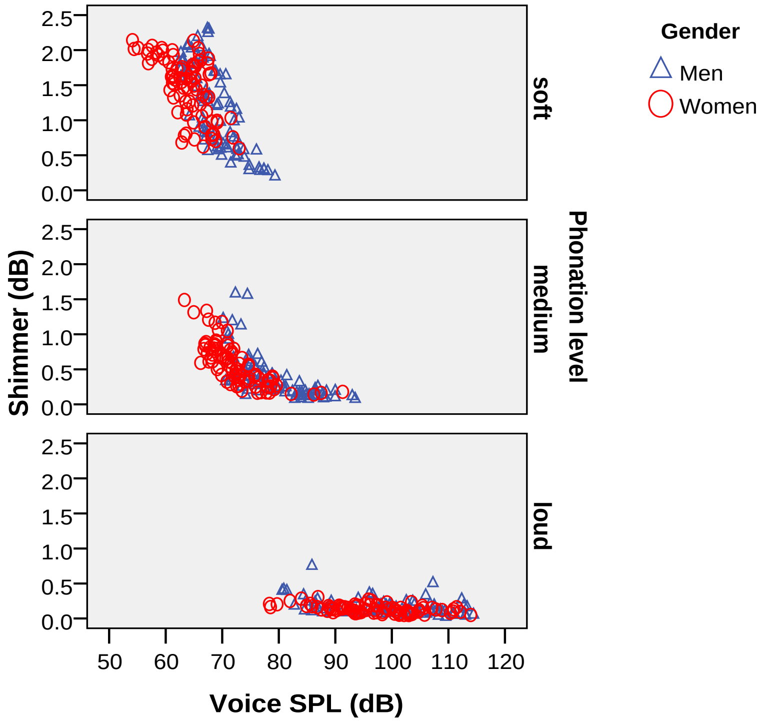
4.4 Results

4.4.1 *Voice intensity effects*

In both genders, the phonation level (soft/medium/loud) had a highly significant effect on jitter and shimmer ($p<0.01$). Jitter and shimmer increased considerably with decreasing voice SPL, especially in phonations below a critical voice SPL of 80dBA (Figure 20). Figure 20 illustrates the dramatic relation of shimmer (dB) with voice SPL (dBA) according to phonations at “soft”, “medium” and “loud” individual voice intensity in women and men.

Mean voice SPL was significantly different in soft, medium and loud phonation (Figure 20, Figure 21, $p<0.01$). Figure 21 shows the mean voice SPL with 95% Confidence Intervals (CI) for both women and men according to “soft”, ”medium” and “loud” voice. Also, F_0 increased significantly from medium to loud phonations ($p<0.01$). Therefore, also voice F_0 might be a clinically relevant influencing factor.

Figure 20: Relation between voice SPL and shimmer



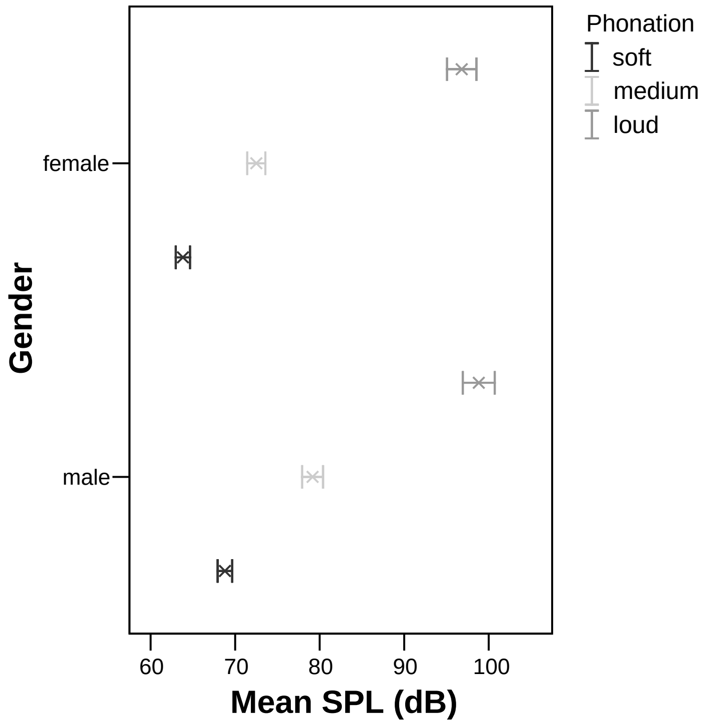
Relation between voice SPL (dBA) in “soft”, “medium” and “loud” individual voice loudness and shimmer (dB) in 28 women and 29 men phonating a prolonged /a/. Origin: own graph.

4.4.2 *Gender effects*

In soft and medium phonations, men had a significantly higher voice SPL than women ($p<0.01$). As expected, there was a significant difference in fundamental frequency (F_0) between women and men ($p<0.01$).

Men had significantly less shimmer than women in soft and medium phonations ($p<0.01$). However, jitter was significantly higher in men when phonating with a soft voice ($p<0.01$). The effects of phonation level (soft/medium/loud) were significantly different in women and men for shimmer ($p<0.01$), but not for jitter.

Figure 21: Mean SPL and 95% CI for soft, medium and loud voice in women and men



Mean voice SPL (dBA) with 95% Confidence Intervals (CI) in “soft”, “medium” and “loud” phonation as measured in 28 women and 29 men phonating a prolonged /a/. Origin: Brockmann et al., 2008.

4.5 **Discussion and conclusions**

This pragmatic study showed a dramatic influence of the individual`s speaking voice SPL on shimmer. A similar, but weaker effect was found for jitter (Appendix B). Also, gender had a significant impact on acoustic voice perturbation (Brockmann et al., 2008). Especially in phonations below 80dBA at "normal" and "soft" individual loudness small SPL variations led to large jitter and shimmer changes (Figure 20).

Men phonated significantly louder than women at "normal" voice loudness. Thus, the significant gender effects might be mainly linked to different habitual voice SPL at individually "normal" voice loudness (Figure 20). In clinical assessments, requesting phonations above 80dBA at a comparable loudness between genders might enhance measurement reliability. However, gender effects might be also linked to natural differences in fundamental frequency (F_0) between women and men.

To determine which of the factors, voice SPL, gender or $F_{0,}$ should be controlled in voice clinics, their relative influence has to be understood. This will be addressed in the first study of the present thesis (chapter 5).

4.6 **Summary of findings**

- In this study it was shown for the first time that differences in habitual voice SPL at individually "normal" voice intensity have a dramatic influence on both jitter and shimmer.

- The observed significant gender effects might be linked to systematically different voice SPL and F_0 between women and men.

- To determine if voice SPL, F_0 and gender should be controlled for in clinical measurements, their relative importance has to be determined.

Chapter 5. The relevance of voice intensity, fundamental frequency, gender and vowel effects in a typical clinical voice task

The main body of the work in this present chapter 5 has been published as follows:

Brockmann, M., Drinnan, M. J., Storck, C. and Carding, P. N. (2011) 'Reliable Jitter and Shimmer Measurements in Voice Clinics: The Relevance of Vowel, Gender, Vocal Intensity, and Fundamental Frequency Effects in a Typical Clinical Task.', Journal of Voice, 25, (1), pp. 44-53.

Healthy adults phonate with an unequal voice SPL when performing the commonly applied voice task to say a vowel at “comfortable loudness”. The preceding chapter 4 showed that differences in voice SPL have a dramatic influence on both jitter and shimmer. Also gender has a significant impact on voice perturbation, which might be linked to systematically different voice SPL, but also unequal F_0 between women and men. Furthermore, from chapter 2 we are aware that different vowels might be a further considerable confounding factor in clinical jitter and shimmer measurements. In the present chapter the relative importance of voice SPL, F_0, gender and vowels and their interactions was assessed to understand how these factors might be controlled for in voice assessments.

5.1 Introduction

In clinical assessments patients are usually instructed to sustain the vowels /a/, /u/, /o/ or /i/ “at comfortable loudness and pitch” (Dejonckere et al., 2001). This voice task is used to avoid potentially confounding intensity and pitch change effects on acoustic measurements as well as a possible influence of articulatory movements during speech (Dejonckere et al., 2001; Titze, 1995; Heiberger and Horii, 1982). However, it is unclear whether these instructions are sufficiently rigorous to produce maximum reliability in clinical practice.

A review of the current literature (chapter 2) shows that the confounding effects of voice SPL, F_0, vowels and gender have not been satisfactorily described to date. This might contribute to the limited reliability of jitter and shimmer measurements in voice clinics (Carding et al., 2004; Zyski et al., 1984). Therefore, the influence of each factor has to be evaluated in a voice task modelling clinical practice.

In the preceding chapter 4, the effects of voice SPL and gender on jitter and shimmer were assessed (Appendix B). Jitter and shimmer dramatically increase and show far more spread below a critical voice SPL of 80dBA, when adults are asked to phonate at “soft”, “normal” and “loud” individual vocal intensity (Appendix B, Brockmann et al., 2008). From previous work by Brown et al., it is clear that when healthy adults are asked to phonate “at comfortable loudness and pitch”, voice SPL varies considerably between individuals. Across several speaking conditions a mean voice SPL of 63 to 75dBA has been reported (Brown et al., 1996). Thus, “comfortable” voice intensity of both women and men always lies below the above described threshold of 80dBA, which may partially account for the considerable jitter and shimmer differences between individuals.

Also, women phonate systematically softer in the same voice task than men (Brockmann et al., 2008; Brown et al., 1996). Therefore, habitual differences in voice SPL between women and men might underlie the observed gender effects (Brockmann et al., 2008). In addition to this, if different vowels are naturally phonated at different vocal intensity, then voice SPL would become a major confounding factor underlying both vowel and gender effects.

Further, natural and systematic differences in fundamental frequency (F_0) between women and men might contribute to gender effects (Gelfer, 1995; Sussmann and Sapienza, 1994). In studies investigating the gender influence it was often concluded that higher F_0 is associated with higher jitter (Fitch, 1990; Nittrouer et al., 1990). However, when women and men changed their F_0, jitter and shimmer were highest in lower frequencies (Gelfer, 1995; Orlikoff and Baken, 1990). Furthermore, when women and men are asked to lower their F_0, men showed a stronger decrease of jitter and shimmer than women (Orlikoff and Baken, 1990). As already described by Baken, it can be concluded that the influence of F_0 changes has not been sufficiently understood to date (Baken and Orlikoff, 2000b; Baken and Orlikoff, 2000a).

Several studies investigating vowel effects in healthy adults described significant jitter and shimmer differences between vowels. However, in these studies F_0 also (significantly) changed from vowel to vowel (Dwire and McCauley, 1995; Fitch, 1990; Nittrouer et al., 1990; Linville and Korabic, 1987; Wilcox and Horii, 1980). Further, Orlikoff found no variation in jitter and shimmer with vowel when F_0 and SPL were controlled for (Orlikoff, 1995). None of the described studies investigated the interaction between F_0 and vowel in a typical clinical voice task.

Therefore, based on the current literature and the results of the preceding MSc study (chapter 4, Appendix B) we have to expect that an influence of F_0 and voice SPL might underlie the reported gender and vowel effects (Brockmann et al., 2008). This is especially true in a clinical voice task where combined effects between these variables might emerge. Therefore, in the present study the influence of gender and vowel are assessed in a typical clinical voice task, while the confounding effects of voice SPL and F_0 are corrected for. Further, the relative influence of each factor will be investigated. This might help to derive recommendations for improving clinical measurements of jitter and shimmer.

5.2 **Specific study aims**

In this chapter the following study questions will be investigated in vocally healthy adults:

(1) Do vowel and gender have an effect on jitter and shimmer in a typical clinical voice task when the confounding effects of voice SPL and F_0 are corrected for?
(2) How big are the effects due to vowel, gender, SPL and F_0 on jitter and shimmer?
(3) How can we control for the described effects in clinical measurements of jitter and shimmer?

5.3 **Methods**

In the present chapter, the data body described in chapter 4 was further analysed (Brockmann et al., 2008). Please refer to chapters 3 and 4 and Appendix B for further details on the methods used.

In a cross-sectional single cohort study 57 vocally healthy volunteers (28 women, 29 men) were recorded while phonating the vowels /a/, /o/ and /i/ at individually “soft”, “normal” and “loud” voice in randomised order. Each intensity level (soft/normal/loud) was repeated three times, giving a total of 1539 single phonations for acoustic analysis. To model the typical clinical voice task “please phonate at comfortable loudness and pitch” only the phonations at “normal” voice intensity with the vowels /a/, /o/ and /i/ (i.e. 513 single phonations) were used for analysis.

5.3.1 ***Participant inclusion criteria***

Please refer to chapter 3.

5.3.2 ***Voice recording technique***

Please refer to chapter 3 for details on the voice recording technique.

5.3.3 ***Vowel choice and voice recording protocol***

<u>Vowel choice</u>
In the present study the vowels /a/, /o/ and /i/ were chosen to ensure that all participants produce comparable phonations. In the literature, mostly the vowels /a/, /i/ and /u/ have been used to investigate vowel effects (please refer to section 2.5.4 for a detailed discussion). However, the studies presented in chapter 4 and 5 were conducted in Switzerland, and the participants spoke one of the three national languages Swiss German, Italian and French. In pretests it was observed, that the vowel /u/ could not be spontaneously produced by all French speaking participants. Also, the Swiss German speakers articulated /u/ generally markedly backwards. This was not the case for the vowels /a/, /o/ and /i/.

Voice recording procedure

All participants were allowed to practice the given voice tasks for a maximum time of 10 minutes. First, the participants were asked to “sustain /a/ for 5 seconds at habitual pitch and loudness”. When able to do this, they were asked to “sustain /a/ for 5 seconds, as softly as possible” and then “as loudly as possible”. The identical procedure was repeated with the vowels /o/ and /i/. These instructions were chosen to assess voice variance in uninfluenced voice function as done similarly in clinical practice.

When the vowels were not correctly articulated, the examiner gave feedback and said example words including the target vowel in High German. For all voice tasks the examiner used verbal instructions only and did not model.

When the participants felt able to produce the requested loudness levels and correctly articulated the vowels /a/, /o/ and /i/, voice recordings were made in randomised order. Each participant provided three repetitions of each vowel and loudness level, giving 27 recordings (3 of each vowel /a/, /o/, /i/ in three loudness levels). This gave a total pool of 1539 single vowel phonations. Of these, only 513 recordings at “normal” voice intensity with all vowels /a/, /o/ and /i/ were used in the present study.

5.3.4 *Main outcome measures*

In the present study the instrumental acoustic analysis was conducted with PRAAT (Boersma and Weenink, 2006). The sample used was 0.5 to 3.5 seconds from voice onset of each phonation.

The main outcome measures were:

- jitter (%) called “Jitter (local)” in PRAAT,
- shimmer (dB) called “Shimmer (local dB)”,
- fundamental frequency (Hz) called “mean pitch”,
- voice SPL (dB) called “mean energy intensity”.

Calibrated voice SPL dBA was calculated by comparison method, as described in detail in chapter 3 (Winholtz and Titze, 1997a).

5.3.5 *Statistical analysis*

From the preceding study to assess the effects of voice SPL and gender we were aware that jitter and shimmer both showed a log-normal distribution (Appendix B; Brockmann et al., 2008). Therefore, a logarithmic transform was applied to each before further statistical analysis.

(1) Do vowel and gender have an effect on jitter and shimmer in a typical clinical voice task when the confounding effects of voice SPL and F_0 are corrected for?

First, analysis of covariance (ANCOVA) was used to assess the effects of the continuous explanatory factors F_0 and voice SPL and the fixed factors vowel (/a/, /o/, /i/) and gender (f/m). Also, the first-order interactions between all four factors were assessed; for example this would indicate whether vowel effects are different in women and men.

Subjects (1-57) were included as a random factor within gender. Thus, it was assumed that the male and female subjects are simply random samples of the population being studied. The three repeats (1-3) of each experiment, which ideally would give identical jitter (or shimmer) measurements, were used to estimate the random (or unexplained) measurement variance not explained by the factors investigated in this study.

(2) How big are the effects due to vowel, gender, SPL and F_0 on jitter and shimmer?

Finally, the effect size of each factor vowel, gender, SPL and F_0 were assessed using the eta squared statistic. Eta squared estimates the proportion of the overall variance in jitter (or shimmer) that can be attributed to each of the investigated factors. Each factor has its own eta-squared value, and the sum of all eta-squared values (including the measurement error) is 1. Thus, eta-squared gives an indication of which factors affect jitter (or shimmer) the most, and are therefore the most important.

(3) How can we control for the described effects in clinical measurements of jitter and shimmer?

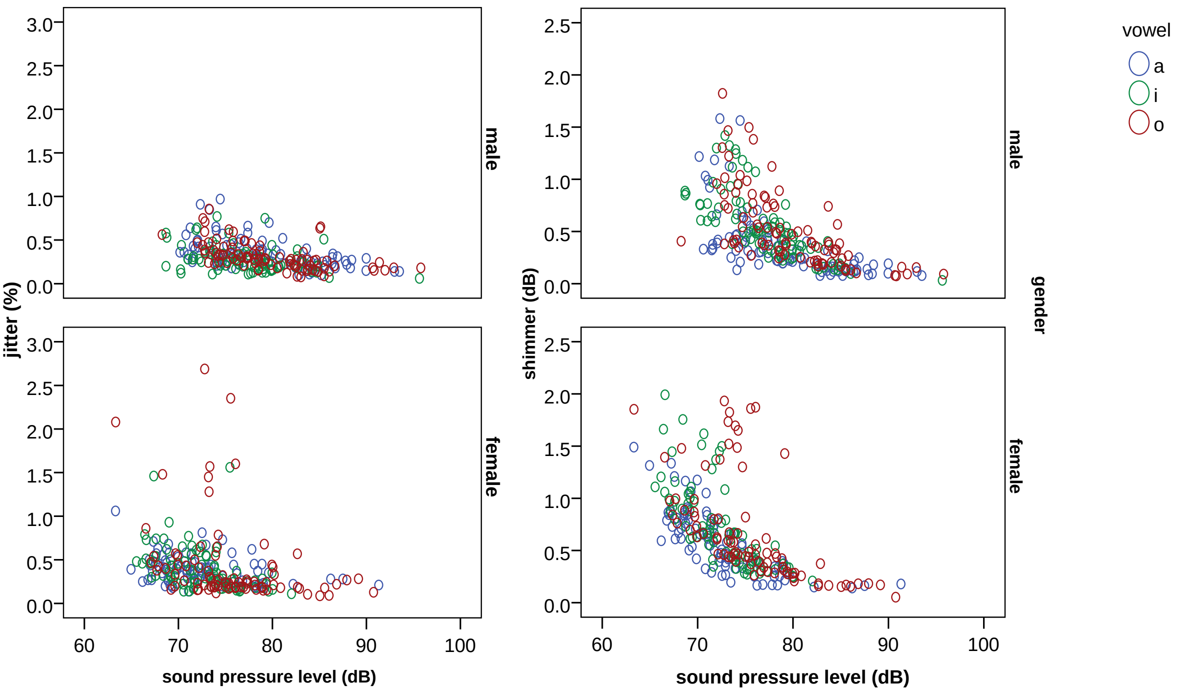
To address this question, modifications to the usual clinical voice assessment protocol were proposed in the discussion (section 5.5.4). Further, the potential improvements were evaluated by calculating the relative sensitivities of the original and the revised assessment protocol (section 5.5.5). This was done by removing the measurement variance caused by vowel and voice SPL, as given by the eta-squared statistic. Further, to estimate the effect of 6 repetitions, the random error was reduced by the factor of square root of 6, approximately 2.4.

5.4 Results

5.4.1 *Do vowel and gender have an effect on jitter and shimmer in a typical clinical voice task when the confounding effects of voice SPL and F_0 are corrected for?*

In Figure 22 the effects of voice SPL, gender and vowel on jitter and shimmer are shown. Even without resort to statistics, it is clear that voice SPL has a dramatic confounding effect on jitter and particularly shimmer (Figure 22). The effects of gender and vowel are not so apparent and seem to mostly influence jitter and shimmer measurements in phonations below 80dBA. The effects of F_0 are not considered in Figure 22.

Figure 22: Relation of jitter and shimmer with gender, vowel and voice SPL



The relation of jitter (%) (left side) and shimmer (dB) (right side) with gender, vowel and voice SPL (dBA). Clearly voice SPL represents the single biggest source of variability in the measurements. The effect of fundamental frequency is not considered in this figure. Origin: Brockmann et al., 2011.

Table 6 summarises the descriptive statistics for all acoustic variables in men and women. Given are the mean values and 95% Confidence Intervals (CI) for the parameters voice SPL (dBA), F_0 (Hz), jitter (%) and shimmer (dB). Please note that the Confidence Intervals are not symmetrical for women and men, this is a consequence of the logarithmic transform.

In every case the men phonated systematically louder than the women with lower fundamental frequency (F_0). Equally, the men always had lower jitter and shimmer values (more periodic voice signals). It is also evident from Figure 22 that women are systematically quieter in their phonation than men (Table 6).

Table 6: Mean values and 95% CI for voice SPL, F_0, jitter and shimmer according to vowel and gender.

Instrumental parameter	**Men**				**Women**		
	/a/	**/i/**	**/o/**		**/a/**	**/i/**	**/o/**
Voice SPL (dBA)	78.5 (77.6 to 79.3)	77.5 (76.6 to 78.4)	80.2 (79.4 to 81.1)		73.2 (72.3 to 74.1)	72.3 (71.4 to 73.2)	75.0 (74.1 to 75.9)
F_0 (Hz)	127 (123 to 132)	143 (139 to 148)	129 (124 to 133)		214 (209 to 218)	230 (225 to 234)	215 (211 to 220)
Jitter (%)	0.30 (0.28 to 0.33)	0.26 (0.24 to 0.29)	0.27 (0.24 to 0.29)		0.37 (0.33 to 0.40)	0.32 (0.29 to 0.35)	0.32 (0.29 to 0.35)
Shimmer (dB)	0.31 (0.27 to 0.35)	0.44 (0.39 to 0.49)	0.38 (0.34 to 0.43)		0.46 (0.41 to 0.51)	0.65 (0.58 to 0.73)	0.57 (0.50 to 0.64)

Mean values and 95% Confidence Intervals (CI) for each acoustic parameter with respect to vowel and gender. Please note that the confidence intervals for jitter and shimmer are not symmetrical; this is a consequence of the logarithmic transformation. Origin: Brockmann et al., 2011.

Table 7 summarises the results of the statistical assessment by ANCOVA; this model was used to investigate the confounding effects of the continuous variables voice SPL and $F_{0,}$ and the fixed factors of gender and vowel. For each effect, the degrees of freedom (d.f.), the p-value, and eta-squared, which represents the proportion of the total variance explained by the effect in question, are reported. The effects marked with an * are considered as statistically significant. Furthermore, random measurement variance, the variability not explained by the present investigated factors, was included in the table.

All of the factors voice SPL, F_0, gender and vowel have a *statistically significant* effect on either jitter, or on shimmer, or both. For the interactions, there is no clear picture with some significant effects. Nevertheless, statistical significance simply indicates that an effect is probably not attributable to chance alone. However, it gives no indication of how big the effect is, and therefore, we must consider separately which of these effects are clinically relevant.

Table 7: Relevance of the factors voice SPL, F_0, gender, vowel and subject as given by ANCOVA

Factor		**Jitter**		**Shimmer**	
	d.f.	**p**	**eta sq.**	**p**	**eta sq.**
		Main effects			
Voice SPL	1	< 0.001*	0.24	< 0.001*	0.62
F_0	1	0.02*	0.03	0.02*	0.02
Gender	1	0.002*	0.04	0.4	0.00
Vowel	2	0.1	0.00	< 0.001*	0.06
Subject	55	< 0.001*	0.33	< 0.001*	0.18
		Interactions			
Gender with voice SPL	1	0.8	0.00	0.7	0.00
Gender with F_0	2	0.01*	0.01	0.05*	0.00
Vowel with voice SPL	1	<0.001*	0.01	0.9	0.00
Vowel with F_0	2	0.3	0.01	<0.001*	0.00
Gender with vowel	2	0.8	0.00	0.06	0.00
		Random measurement variance not explained by the model			
ERROR			0.34		0.11

Results of the ANCOVA model. For each effect are given the degrees of freedom (d.f.), the p-value, and eta-squared, which represents the proportion of the total variance explained by the effect in question. The effects marked with an * are considered as statistically significant. Origin: Brockmann et al., 2011.

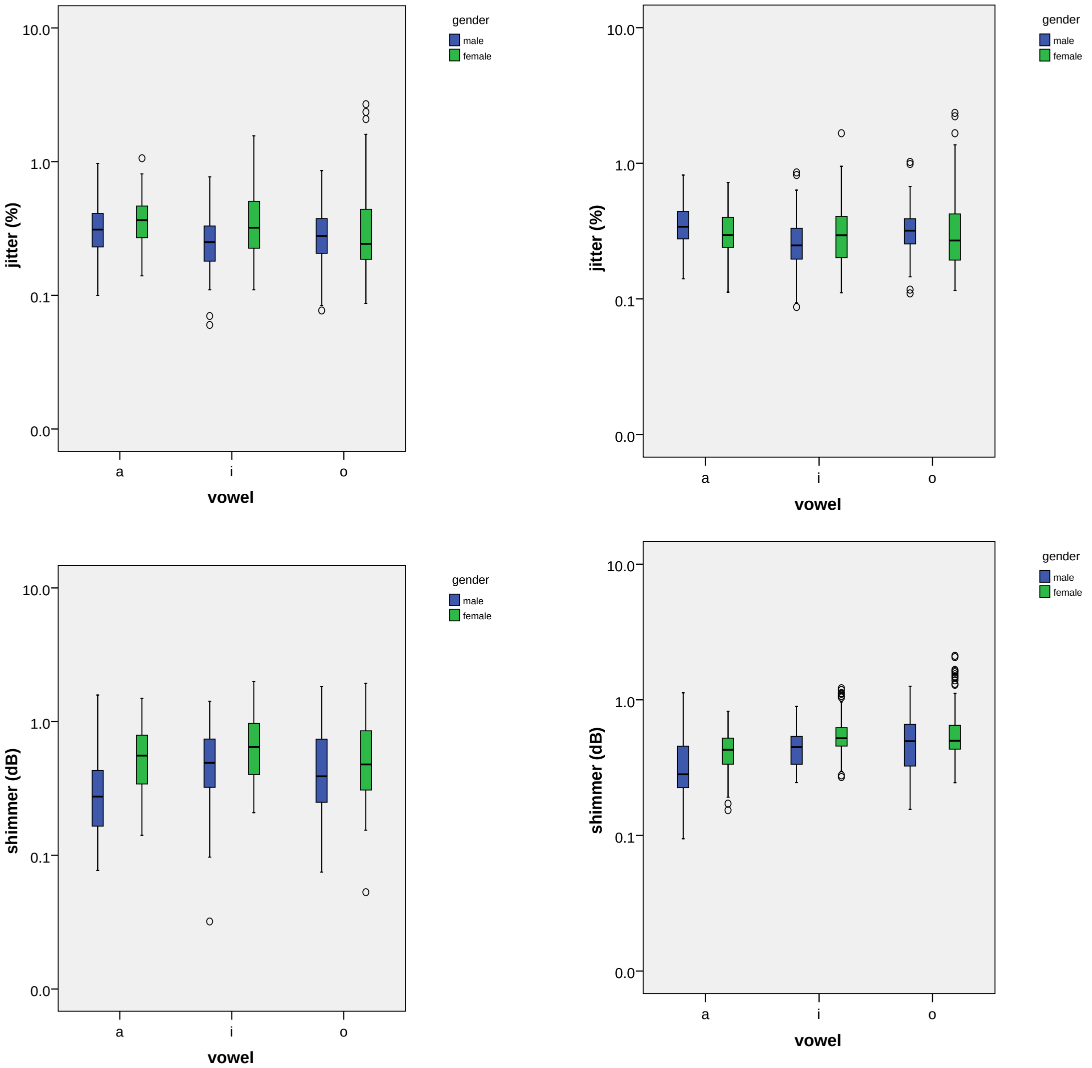
5.4.2 *The effect of correcting for the confounding factors*

The box and whisker plots in Figure 23 show the median, interquartile range, 95% range, and outliers for (top) jitter and (bottom) shimmer. The left panel displays the original jitter and shimmer measurements by gender and vowel, as summarised in Table 7. The graphs on the right side show the effect of correcting each jitter and shimmer measurement according to its own voice SPL and F_0 by ANCOVA.

There are a number of key observations to be made from Figure 23, moving from (*left*) uncorrected to (*right*) corrected measurements. First, there is less spread in the measurements, because the variability introduced by voice SPL and F_0 has been removed. Second, in every case the gap between the genders has changed in favour of women appearing to have relatively less irregular voice signals. For jitter, after correction the difference has gotten smaller but is still statistically significant. For shimmer, there is no evidence of any difference between the genders.

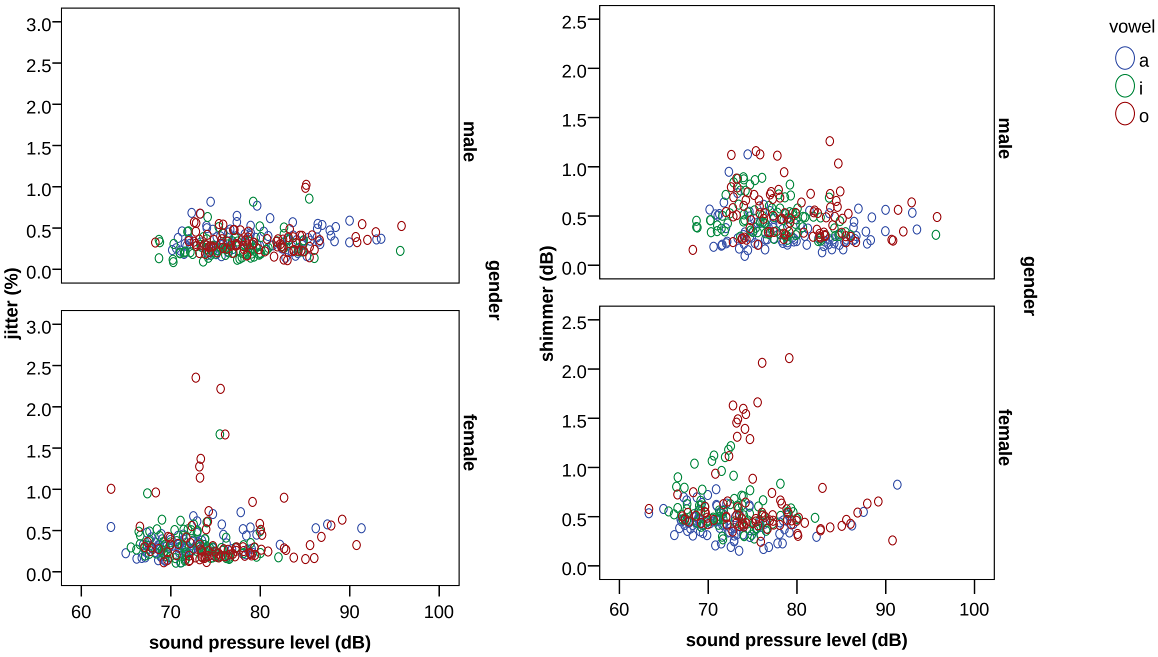
Finally, Figure 23 is directly comparable with Figure 22, showing the results after the effects of voice SPL have been modelled by ANCOVA and subtracted from the original data. For comparison, the measurements are shown against their original voice SPL, thus, the distribution on the x-axis is identical. In both jitter and especially shimmer, the spread against the y-axis is considerably less and shows no relation with F_0.

Figure 23: Median, interquartile range, 95% range, and outliers for jitter and shimmer



Box and whisker graphs showing the median, interquartile range, 95% range, and outliers for (top) jitter and (bottom) shimmer. In the left panel are the original measurements as summarised in Table 6. In the right panel are the measurements after correction for voice SPL and F_0 using ANCOVA. Origin: Brockmann et al., 2011.

Figure 24: Relation of jitter and shimmer with gender and vowel after correction for voice SPL

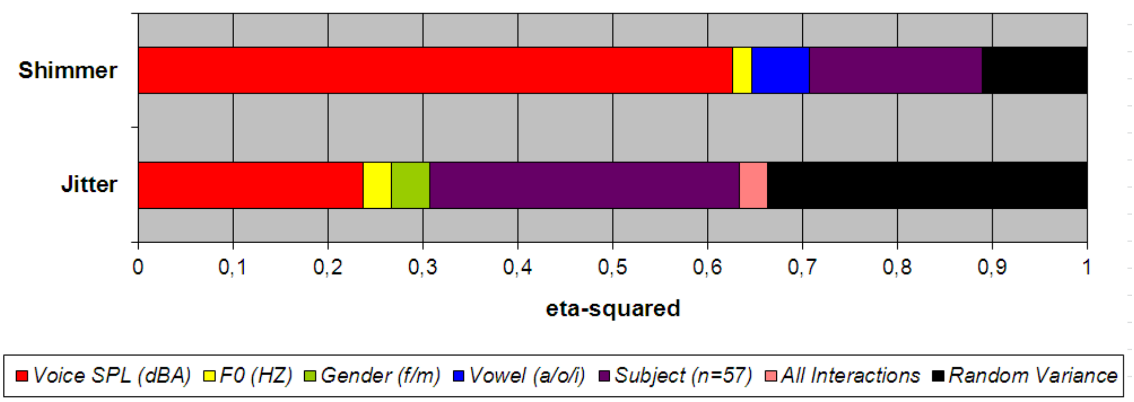


Relation of jitter (%) (left side) and shimmer (dB) (right side) with gender and vowel, after the effect of voice SPL has been modelled using ANCOVA, and subtracted from the original data. For comparison with Figure 22, the measurements are shown against their original voice SPL (i.e., the distribution on the x-axis is identical). However, the spread against the y-axis is considerably less and has no relationship with F_0. Origin: Brockmann et al., 2011.

5.4.3 *How big are the effects due to vowel, gender, SPL and F_0 on jitter and shimmer?*

Figure 25 illustrates the eta-squared measures of effect size taken from Table 7. These sum to 1 and give an estimation of the relative contributions of the various sources of measurement error. In each case, the voice SPL, the inter-subject differences and random (unexplained) measurement variance are clearly the biggest sources of variability in measurements of jitter and shimmer. By comparison, the combined effects of F_0, gender, vowel and the interactions between all factors are relatively small.

Figure 25: Relative influence of voice SPL, F_0, gender, vowel and random measurement variance on jitter and shimmer



The sources of variance for jitter and shimmer, as given by the eta-squared statistic. It is clear that in each case, the voice SPL, intersubject differences, and random (unexplained) measurement variance are the biggest sources of variability. Eta-squared always sums to 1. Origin: Brockmann et al., 2011.

5.5 Discussion

5.5.1 *Do vowel and gender have an effect on jitter and shimmer, when the confounding effects of vocal intensity and fundamental frequency are corrected for?*

In the present study all factors examined, vowel, gender, voice SPL and F_0 and a number of interactions between them showed measurable and statistically significant effects on either jitter, or shimmer, or both (Table 7). Regardless of vowel men always phonated significantly louder and at a lower F_0 than women. Correcting for voice SPL and F_0 substantially reduced the overall measurement variability, which was particularly evident for the shimmer measurements (Figure 23).

Despite the correction for voice SPL and F_0 vowel significantly influenced shimmer. Therefore, vowel effects must be considered as an independent influencing factor in shimmer measurements. Surprisingly, gender was not. After the large correction for voice SPL, there was no evidence that there is any difference between men's and women's shimmer measurements (Figure 23). In contrast to previous findings, women appear not to have naturally higher shimmer when the confounding effects of voice SPL are corrected for (Dwire and McCauley, 1995; Deem et al., 1989; Sorensen and Horii, 1983). The earlier results might be explained by the fact that women phonate systematically softer in the same clinical voice task (Brown et al., 1996), which was also the case in the present study.

For jitter, the subject's gender had a statistically significant effect, but vowel did not. Thus, in principle, jitter measurements do not need to be corrected for vowel differences, but for gender. Notably, in the present study the correction of SPL and F_0 led to different gender and vowel effects on jitter and shimmer. However, as described in chapter 2, we would expect jitter and shimmer to covary and thus to behave in a comparable way. This illustrates that jitter and shimmer might measure different physiologic or pathologic voice features, when the confounding factors are adequately controlled for.

Further, this shows that the interaction between the influencing factors voice SPL, F_0 and gender might be far more complex than previously assumed. For example, in this study women had higher jitter in some vowels that cannot be explained by their softer phonation and higher fundamental frequency. This supports the idea of the physical linkage hypothesis (section 2.5.4), that vowel effects might be produced by an interrelation between articulatory movements in the upper vocal tract and laryngeal muscle activity (Lin et al., 2000; Honda, 1983). As discussed in detail in section 2.5.4, in this case jitter and shimmer would indirectly indicate changes in vocal fold closure and/or tone.

In principle, all influencing factors voice SPL, F_0, gender and vowel could be measured in the clinic, and the jitter and shimmer measurements could be adjusted accordingly. Indeed, some clinical measurement systems have separate normative values for men and women. However, a complex correction with multiple factors clearly becomes unwieldy and unlikely to be used in practice. Further, this might not entirely control for the underlying interaction between the single influencing factors or for physiologic changes due to vowel. Therefore, in the present study eta-squared was used to assess which effects are clinically important in the context of the other factors known to affect acoustic measurements.

5.5.2 *How big are the relative effect sizes of vowel, gender, SPL and F_0 on jitter and shimmer?*

Eta squared gives an estimation of the proportion of the overall variance in jitter or shimmer that can be explained by a factor. By convention, a value of 0.01 or lower is considered a small (and neglible) effect. Therefore, by reference to Table 7 and Figure 25, it can be concluded that all the interaction terms can be neglected. This is fortunate; for example, a large interaction between gender and vowel would mean that a separate correction for each combination of gender and vowel should be used. However, despite a low influence as given by eta-squared there still might be clinically relevant effects. For example we have to be aware that women and men have systematically different F_0, which might also underlie the observed gender effects. Further, in the present study the participants automatically increased F_0 with SPL (Appendix B). This has been described as a physiologic reaction associated with increased vocal fold tone in loud phonations (Gramming, 1988). Therefore, changes in F_0 might underlie the observed gender or SPL effects. This will be addressed in the next study (chapter 6).

Moving to the main effects, it is clear even from Figure 22 that voice SPL has a dramatic effect on both jitter (eta-squared = 0.24) and particularly shimmer (eta-squared = 0.62). That is, almost two thirds of all the variance in acoustic shimmer measurements is simply due to variability in voice sound pressure level. Fundamental frequency plays a considerably smaller though statistically significant role (eta-squared = 0.02). Therefore, it can be concluded that voice SPL differences have to be considered the predominant confounding factor in clinical measurements of jitter and shimmer (Brockmann et al., 2008; Brown et al., 1996). Given this, both jitter and shimmer may be far more robust clinical measures when voice SPL is adequately controlled for.

5.5.3 *How important are these effects in the context of between-subjects effects and random measurement variance?*

In clinical diagnostics, we are most interested in the variability between subjects (i.e. the difference between one subject and the next, who have different clinical conditions). If we know the between-subject variability, then any effects that are small by comparison are arguably unimportant.

As already noted, voice SPL is the single most important influencing factor in the present study. For shimmer, the effect of voice SPL is greater even than the variability due to subject (between-subject variability). The other factors F_0, gender and vowel are considerably less important. In particular, while it seems intuitively almost essential to control for gender in acoustic measurements, the evidence is much less clear. Certainly, the present data suggests that there is just as much reason to control for vowel as for gender.

By taking multiple measurements (3 repeats) from each subject, the random measurement variance due to factors that were not controlled for in this experiment was estimated. This might include variability due to differences in the participant`s performance from one vowel to the next or measurement error. Thus, an indication of the typical variance on a single measurement is given. In the present study random measurement variance was relatively large and similar to the between-subjects variability (Figure 25). This illustrates the fundamental uncertainty of instrumental acoustic measurements in voice clinics, and that we are far from understanding all influencing factors. Therefore, random measurement variance always has to be considered when interpreting acoustic assessment results.

5.5.4 *How can we control for the described effects in clinical measurements of jitter and shimmer?*

From a statistical perspective the effects of voice SPL and F_0 can be relatively easily measured and controlled for; Figure 24 shows the effect of this correction. However, this raises a number of methodological problems. Firstly, results of the present study may not be generalizable to clinical analysis tools that use different algorithms to measure jitter and shimmer. Also, doing so would require the co-operation of the equipment manufacturer to measure voice SPL and perform the correction. Further, it is not clear if this would satisfactorily control for the underlying interrelation between all influencing factors voice SPL, F_0, vowel and gender. As discussed in chapter 2, there might be a complex physiologic interaction based on coordinated movements of articulatory and phonatory muscles. A change in vocal fold tone might indirectly influence the vibratory properties of the vocal folds, und hence also jitter and shimmer.

Therefore, in the following study in chapter 6, a variety of voice tasks potentially suitable for an efficient control of voice SPL and F_0 effects will be investigated.

In the meantime, since gender has such a clear effect on voice SPL, gender-specific norm values should be used where possible. These normative values are currently built into many, but not all, acoustic analysis systems. Given that voice SPL has such a dramatic effect on both jitter and shimmer, we also suggest that women and men should phonate at a standard voice SPL in acoustic assessments. Based on the evidence from our earlier work a standard voice SPL at a minimum of 80dBA (at 10cm distance) should be used (Brockmann et al., 2008). However, it has to be further investigated if it introduces jitter or shimmer when adults or voice patients are asked to produce a target voice SPL, or if patients in voice clinics are able to match this intensity level satisfactorily. Based on current evidence from works by Orlikoff (Orlikoff, 1995) and Gelfer (Gelfer, 1995) we know that jitter and shimmer should not increase considerably due to this change to the usual clinical voice task. Furthermore, 80dBA was achievable even for patients with functional dysphonia (Gramming, 1988) or polyps (Pabon and Plomp, 1988).

Likewise, using different normative values for each vowel would improve the reliability of jitter and shimmer measurements. However, this still leaves the difficulty of interpreting clinical assessments and research reports where different vowels are used. A much simpler solution for clinical purposes would be always to use the same vowel in instrumental acoustic assessments. As discussed above, vowel effects might be produced by an interrelation between muscle movements in the upper vocal tract and the larynx. Based on this and from a pragmatic standpoint, an open vowel that is easy to imitate (i.e. articulation movements should be easily seen and interpreted) irrespective of the native language, linguistic competence or individual health problems (such as hearing disorders) would be optimal. To our knowledge the vowel /a/ would fulfil these criteria best. According to the source-filter theory a change in the nasal resonance due to language or dialect variations should not influence jitter and shimmer measurements (Orlikoff, 1995; Kent, 1993). Also a study by Kiliç et al. comparing vowel effects on jitter and shimmer between Turkish and English native speakers, there were no significant differences between languages (Kiliç et al., 2004). Thus we would expect that acoustic assessment results are transferable between languages, when always the same vowel is used. However, the articulation of the vowel /a/ may vary along the front-back phonetic dimension. To date it is not sufficiently described, if this affects jitter and/or shimmer. Therefore, clinicians should ask the patients to produce the /a/ forward, and check whether they are able to perform this.

Random measurement variability by itself cannot (by definition) be completely controlled, but can normally be reduced by taking the mean of repeated measurements. The mean of N measurements would reduce this term by a factor of $\sqrt{N}$. A previous work by Scherer et al. suggests a number of six repetitions in stable voices and 15 repetitions in unstable phonations (Scherer, 1995).

5.5.5 *What improvements can we hope to see?*

In summary, relatively simple changes to the usual clinical protocol (Dejonckere et al., 2001; Titze, 1995) might considerably improve measurements of jitter and shimmer: we recommend to use phonations at a minimum target voice SPL of 80dBA (at 10cm distance), on the vowel /a/ for all measurements, and to take the mean of six phonations.

To estimate the potential efficiency of these measures Table 7 and Figure 25 give an indication of the improvement we can hope to make. In Table 8, the relative sensitivities of three different acoustic measurement protocols are indicated. The sensitivities were calculated based on the results of the eta-squared statistics (Table 7, Figure 25). For convenience, a sensitivity of 100% was assumed for the original clinical measurement protocol (first column). A sensitivity of 200% would correspond directly to the ability to detect changes two times smaller. We assume that the signal variance remains the same. The measurement variance caused by vowel and voice SPL, as given by the eta-squared statistic, was removed. Further, to estimate the effect of 6 repetitions, the random error was reduced by the factor of square root of 6, approximately 2.4.

Table 8: Potential effects of changes to the usual clinical protocol

	N=1 Without any control	Mean of N=6 Controlling for vowel	Mean of N=6 Controlling for vowel and SPL
Jitter	100%	135%	194%
Shimmer	100%	111%	202%

The relative sensitivities of three different acoustic measurement protocols. For convenience, a sensitivity of 100% was assumed for the first protocol. The data for the third protocol are speculative, based on a reduction in voice SPL variability by half. Origin: Brockmann et al., 2011.

Clearly, there are significant gains to be made from relatively simple changes to the usual clinical practice. Column two indicates the increase in sensitivity by using always the same vowel, which would considerably improve measurements of jitter. Finally, in column three the potential effect of controlling for voice SPL was also considered. Although perfect control of voice SPL does not appear humanly possible, it seems plausible that the effects might be reduced by half. This has an enormous effect on the sensitivity of jitter and especially shimmer measurements. Future research should investigate the efficiency of these measures under clinical assessment conditions; this will be one aim of the next study described in chapter 6.

5.6 Conclusions

With this pragmatic study the influence of gender and vowel on jitter and shimmer in a typical clinical voice task was assessed while correcting for the confounding effects of voice SPL and F_0. Surprisingly, the effects of vocal intensity differences between individuals had a stronger impact on both jitter and shimmer than any other factor. By comparison, fundamental frequency had a relatively small influence. Also, vowel and gender effects were considerably smaller, but statistically significant and clinically important. Therefore, current guidelines are not sufficient to control for these effects in clinical practice.

According to the eta squared statistic the interactions between voice SPL, F_0 and gender had no significant influence on jitter and shimmer. Since F_0 is systematically higher in women and increased with voice SPL in both genders, there still might be a clinically relevant influence contributing to the observed gender and voice SPL effects. This will be addressed in the next study.

Jitter and shimmer might measure different physiologic or pathologic voice features, when the described confounding factors are adequately controlled for. Based on the results from this study, it can be assumed that requesting phonations a predefined and comparable loudness level between women and men, and always using the same vowel, would enhance measurement reliability and sensitivity. For this, we recommend to use the vowel /a/ and phonations at a minimum of 80dBA. Also, gender specific thresholds applying these guidelines should be established. However, the feasibility and efficiency of these measures has to be further verified and tested under clinical assessment conditions. This will be key aim of the study described in chapter 6.

5.7 **Summary of findings**

- In clinical measurements of jitter and shimmer voice SPL is the single biggest influencing factor. Vowel and gender effects had a clinically important impact, whereas fundamental frequency had a relatively small influence.

- Phonations at a predefined voice SPL of 80dBA minimum with the vowel /a/ would enhance measurement reliability and sensitivity. However, the feasibility and efficiency of these measures has to be investigated.

- Gender-specific normative values using these guidelines should be established.

- F_0 is systematically higher in women and naturally increases with voice SPL. Therefore, despite a comparatively low influence as measured by eta squared, F_0 might contribute to the observed gender and voice SPL effects.

Chapter 6. Compliance to voice protocols in healthy adults

The work presented in chapter 6 has been published as follows:

Brockmann, M., Drinnan, M. J., and Carding, P. N. (2009) 'Comparison of voice intensity effects on electroglottographic versus acoustic jitter and shimmer', Journal of the Acoustical Society of America, 125, (4), 2532 (Abstract).

In the previous chapters, I demonstrated that differences in the individual's voice SPL are the single biggest influencing factor in jitter and shimmer measurements. Vowel and gender had a clinically relevant impact, whereas fundamental frequency had comparatively little influence. It was shown that jitter and shimmer reliability could be improved by requesting phonations at a predefined voice SPL of 80dBA minimum with the vowel /a/. In the previous study, F_0 naturally increased with voice SPL in both genders and was systematically higher in women. Despite a comparatively low statistical relevance, F_0 might contribute to the observed gender and voice SPL effects. However, from the literature it is not clear how well healthy adults are able to control for voice SPL and F_0. Further, this change to the usual vocal behaviour might influence jitter and shimmer.

Therefore, in the present chapter untrained healthy women and men performing several voice tasks will be investigated. A variety of predefined voice intensity levels, with and without additional control of F_0 will be tested. Based on this, it will be discussed how we might use this information to derive the best clinical protocol to produce the most useful jitter and shimmer measurements.

6.1 **Introduction**

One main conclusion of the preceding study (chapter 5) is that voice SPL and gender effects in clinical jitter and shimmer measurements might be adequately controlled for when patients phonate at a prescribed voice intensity level of 80dBA minimum (10cm recording distance). Further, from the preceding experiments we are aware that differences in fundamental frequency (F_0) might have contributed to the observed voice SPL and gender effects on both jitter and shimmer (chapter 4 and chapter 5). Mean F_0 increased with voice SPL in the whole investigated group, and women naturally have a systematically higher F_0. Given this, every patient should also phonate at the same predefined F_0 in instrumental acoustic assessments.

However, this raises a number of important pragmatic issues. First, untrained adults might respond in a significantly different way to identical voice tasks, or might even be unable to perform a task. There might also be systematic differences between women and men, which was the case in the preceding study for soft, medium and loud voice. Therefore, to ensure basic measurement comparability, a voice protocol should be used that is doable for the *majority* of patients.

Previous research in jitter and shimmer at prescribed voice SPL and F_0 levels is sparse. An extensive literature review highlighted only two studies on this topic (Gelfer, 1995; Orlikoff, 1995). Gelfer investigated vowel effects in 29 women speaking at their natural F_0 and 1 octave above at the prescribed voice intensity levels of 60, 70 and 80dBA (at 30cm distance). In this study all factors voice SPL, F_0 and vowel had a significant influence on both jitter and shimmer (Gelfer, 1995). However, Orlikoff reported contradictory results. He investigated 10 women and 10 men phonating at a fixed voice SPL of 74dB (at 30cm distance) and prescribed F_0 levels of 220Hz (women) and 110Hz (men) by acoustic and electroglottographic jitter and shimmer (Orlikoff, 1995). In his experiment, no vowel effects were found. Orlikoff concluded that this was due to an adequate control of voice SPL and F_0. From the studies by Gelfer and Orlikoff it is clear that healthy adults are in principle able to produce and maintain prescribed voice intensity levels between 70dBA and 90dBA (at 10cm distance) (Gelfer, 1995; Orlikoff, 1995). Gelfer asked the participants to phonate at 60, 70 and 80dBA as measured at 30cm distance (Gelfer, 1995). According to the inverse-square law this would equal to 70, 80 and 90dBA at 10cm distance as used in the preceding studies of chapters 4 and 5.

However, a transfer of the reported study results into clinical practice is not possible for several reasons. In both studies it was not assessed how well the participants matched the target voice SPL and F_0 levels. Even though Orlikoff reported that several women and men had to repeat the voice tasks “due to an inacceptable frequency level and….an unacceptable vocal SPL” (Orlikoff, 1995), there was no specific information on how accurately the participants performed. Given the large influence of even small voice SPL changes on jitter and shimmer, it is critical how well the participants match the target SPL (chapter 5). Further, control of F_0 is usually difficult for a large proportion of untrained healthy adults. As shown by studies in singing tasks, around 50% of adults are unable to match a predefined F_0 (Estis et al., 2001; Watts et al., 2005). This has implications for the voice tasks we can use in clinical practice. Also, gender effects were not considered in the works by Gelfer and Orlikoff (Gelfer, 1995; Orlikoff, 1995). As shown in the preceding experiment of the present thesis, the participant`s gender is a significant influencing factor in measurements of jitter and shimmer (chapter 5).

These issues will be addressed in the present study by testing a variety of voice tasks in vocally healthy, untrained women and men: in the first task the participants will be asked to produce soft, medium and loud voice; in the second task a variety of voice intensity levels will be tested; in the third task the participants will be asked to produce different predefined voice SPL while keeping their own habitual F_0. Thereafter, the task performance will be assessed separately in women and men.

A further important pragmatic issue might be that a change to the usual voice task could lead to a change in the patient’s vocal behaviour. This could indirectly influence the vibratory properties of the vocal folds and hence also jitter and shimmer (please see also section 2.1). Therefore, in the present study also the effects of different voice task types (task 1: without any control; task 2: control of SPL; task 3: control of SPL and F_0) on jitter and shimmer will be assessed.

6.2 **Specific study aims**

In this chapter the following study questions will be investigated:

(1) How can we improve the clinical protocol in measurements of jitter and shimmer?

(2) Does the voice task type influence jitter and shimmer?

6.3 **Methods**

In this cross-sectional single cohort study 40 vocally healthy volunteers, 20 women (mean age 28;4 years;months) and 20 men (mean age 30;1 years;months) between 20 and 40 years were investigated. All participants were recorded while phonating five seconds of /a/ under 11 voice protocols: at subjective "soft", "normal" and "loud" voice intensity (*task 1*); at the prescribed voice intensities of 65, 75, 85 and 95dBA (with visual feedback, *task 2*); and at the prescribed intensities of 65, 75, 85 and 95dBA and with fixed fundamental frequency (F_0) (with visual and auditory feedback, *task 3*). Table 9 provides a summary of the voice tasks, the recording instructions and how the recordings were labelled.

6.3.1 ***Participant inclusion criteria***

Please refer to chapter 3.

6.3.2 ***Voice recording technique***

Please refer to chapter 3 for details on the voice recording technique, equipment and calibration.

6.3.3 ***Voice recording protocol***

<u>Determining the habitual mean fundamental frequency (F_0)</u>

Prior to testing voice protocols the habitual mean fundamental frequency (F_0) of the participants was determined during normal speech. For this the participants were recorded while answering the question "Please tell me what you did today until now". Thereafter, mean F_0 (Hz) was measured by PRAAT (Boersma and Weenink, 2006) and converted to musical notation (for example: 259Hz = c1). Mean F_0 as determined in this step was later used in voice task 3 to provide auditory feedback (by piano).

<u>Voice task 1</u>

In the first voice task the participants were asked to phonate 5 seconds /a/ "at normal voice loudness and with your habitual pitch", "as soft as possible" and "as loud as possible" (*protocol 1-3*). The participants were allowed to train up to 5 minutes until they were able to produce three auditively discernible different voice intensity levels. Thereafter, they were recorded while phonating /a/ for 5 seconds at each voice intensity level for 2 times (randomised order).

Voice task 2

In the second voice task phonations of 5 seconds of /a/ at the target voice intensities of 65, 75, 85, and 95dBA SPL were required (*protocol 4-7*). The participants could monitor their voice intensity level by visual feedback. For this, the four target voice intensity levels were marked with coloured arrows on a computer screen (please see section 3.2.3). When the participants felt able to match the prescribed loudness levels after 5 minutes of training maximum, they were recorded 2 times at each intensity level (randomised order).

Voice task 3

In the third voice task the participants were asked to phonate 5 seconds /a/ at 65, 75, 85 and 95dBA respectively while keeping their own pitch level at the same time (*protocol 8-11*). Again, the participants were able to monitor their voice intensity on an extra screen (visual feedback, section 3.2.3). Further, additional auditory feedback was provided: before each recording the participant’s individual mean speaking F_0 was played by piano. Also, the participants were specifically trained for this task. First they were asked to phonate /a/ with “comfortable” voice intensity at their own pitch, which was played by piano. Afterwards, they were asked to match the predefined voice intensity levels while keeping their own pitch (each time with visual and auditory feedback). When the participants felt able to keep their pitch at the predefined four voice intensity levels after a maximum of 5 minutes training, 2 phonations of /a/ at each intensity level were recorded (randomised order). Again, before each loudness level the participant´s pitch was played by piano (auditory feedback).

Table 9: Summary of tested voice tasks types and protocols

Voice task type	Protocol number	Label	Instruction by examiner
1	1	Soft	"Please say 5 seconds /a/ at normal loudness/as soft as possible/as loud as possible with your habitual pitch"
1	2	Medium	
1	3	Loud	
2	4	65dB	"Please match the four loudness levels as indicated by the arrows on the screen"
2	5	75dB	
2	6	85dB	
2	7	95dB	
3	8	65dB FF	"Please match the four loudness levels as indicated by the arrows on the screen and keep your pitch level. Your pitch will be played once before the task"
3	9	75dB FF	
3	10	85dB FF	
3	11	95dB FF	

Summary of voice task types (1-3) and protocols (1-11) with oral instructions as used by the examiner. Origin: own table.

6.3.4 *Main outcome measures*

Instrumental acoustic analysis was done by PRAAT using second 2.5 to 4.0 from voice onset of each phonation (Boersma and Weenink, 2006).

The main outcome measures were:

- jitter (%) called "Jitter (local)" in PRAAT,
- shimmer (%) called "Shimmer (local)",
- fundamental frequency (Hz) called "mean pitch",
- voice SPL (dB) called "mean energy intensity".

Calibrated voice SPL dBA was calculated with the comparison method, as described in chapter 3 (Winholtz and Titze, 1997a).

6.3.5 ***Statistical analysis***

A clinical voice recording protocol is only useful, when most patients are able to perform it. To assess how well the participants performed the different tasks, mean voice SPL and F_0 and the Standard Deviation (SD) were determined per voice protocol in women and men.

Since jitter and shimmer again followed approximately exponential distributions with SPL, a logarithmic transform was applied to each before further analysis (please also see chapter 5).

An Analysis of Covariance (ANCOVA) was used to determine the effects of voice task (1-3), gender (f/m) and subject (1-40) on voice SPL, F_0, jitter and shimmer. Since the voice protocol (1-11) depended on the voice task type (1-3), it was included as covariable. To determine if the effects due to voice task were different in women and men, the interaction of voice task and gender was also assessed.

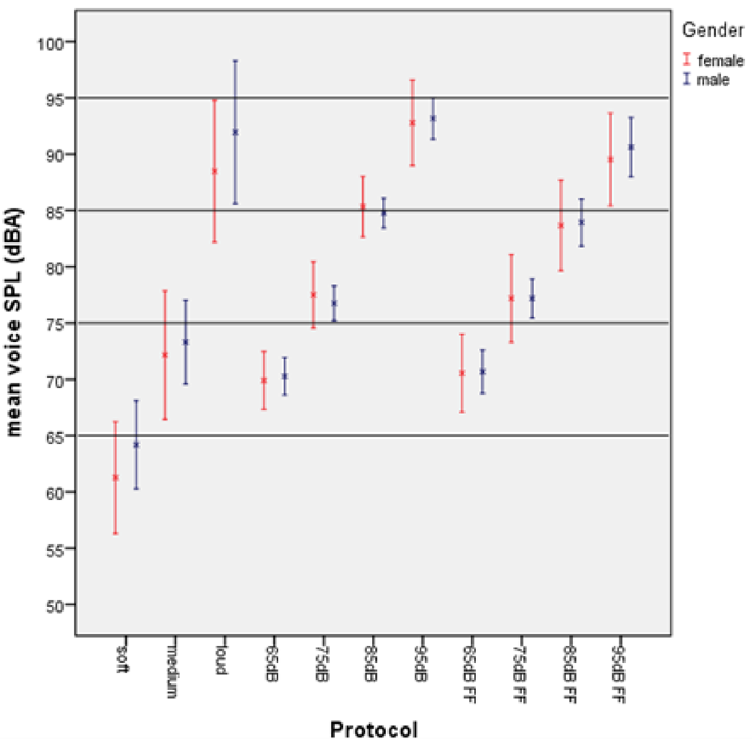
6.4 Results

6.4.1 *How can we improve the clinical protocol in measurements of jitter and shimmer?*

<u>Voice task 2: control of voice SPL</u>

As indicated by Figure 26 the prescribed voice intensity level of 85dBA without control of F_0 was matched best by both women and men (women: 85.3dBA, SD 2.7; men: 84.8dBA, SD 1.3). Also, the Standard Deviation (SD) and thus the spread was lowest in phonations at 85dBA (Figure 26, Table 10). Control of voice SPL was better in voice task 2 (prescribed voice SPL) than in task 3 (prescribed voice SPL with control of F_0). The mean voice SPL varied significantly by voice task and protocol ($p<0.001$, Table 11).

Figure 26: Voice SPL (dBA) per voice task in women and men



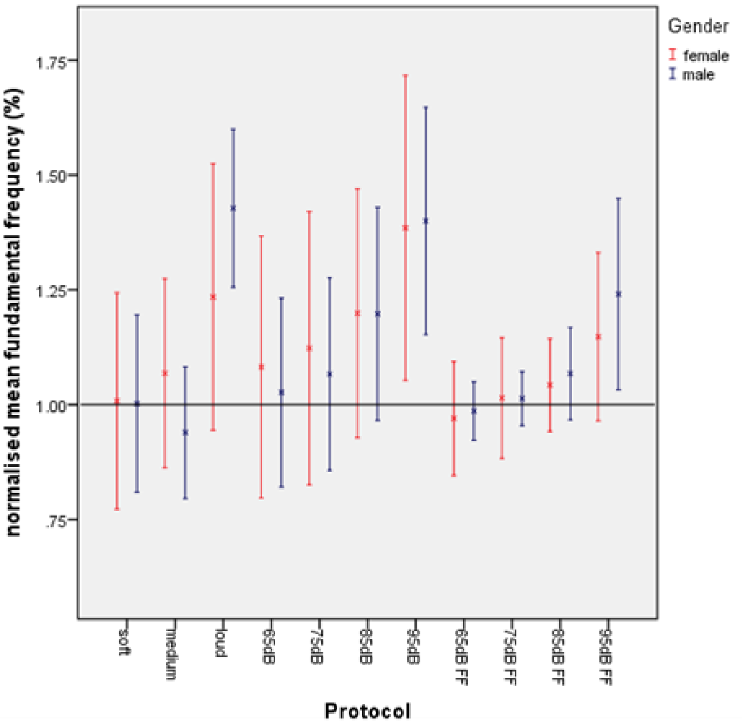
Mean voice SPL (dBA) for women (red) and men (blue) according to the 11 voice protocols. The bars indicate a single Standard Deviation (SD). From the left to the right side are shown: *task 1* individual "soft", "medium" and "loud" phonations, *task 2* prescribed intensity levels of 65, 75, 85 and 95dBA (dB), and *task 3* the prescribed intensity levels with control of fundamental frequency (dB FF). The reference lines indicate 65, 75, 85 and 95dBA. A voice SPL of 85dBA without control of F_0 was matched best by both women and men. Origin: own figure.

Gender differences in mean voice SPL were smaller in voice task 2 and 3 as compared to task 1 (Figure 26, Table 10). However, the effects of voice task on produced voice SPL were significantly different in women and men (p<0.002).

Task 3: control of voice SPL and F_0

Mean F_0 (Hz) increased with higher voice intensities in both women and men regardless of voice task (Table 10, Figure 27). Figure 27 shows the mean F_0 normalised for the individuals` mean speaking F_0. The reference line at 1.00 indicates the 100% level for the overall mean F_0. In women and men the overall mean F_0 was closest to the habitual speaking (target) F_0 in voice task 3 (with control of SPL and F_0) and in "soft" phonations (protocol 1). However, the target F_0 was not exactly matched in voice task 3, especially in phonations at 85dBFF and 95dBFF (Figure 27, Table 10). In agreement with this the voice protocol (1-11) and the task type (1-3) had highly significant effects on mean F_0 in both women and men (p<0.0001, Table 11).

Figure 27: Normalised mean F_0 (%) per voice task in women and men

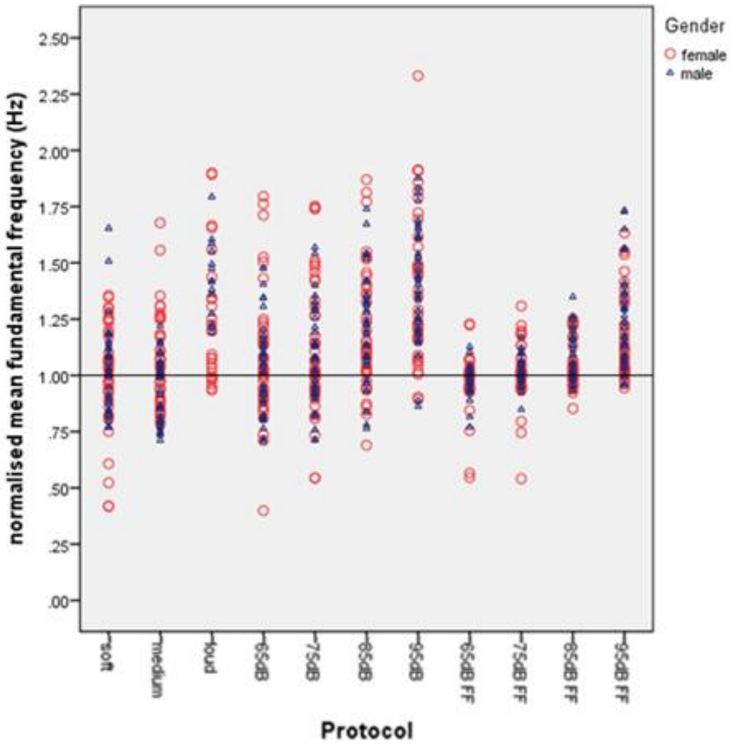


Mean fundamental frequency normalised for the participant`s individual habitual F_0. The reference line at 1.00 indicates the 100% level of the overall target fundamental frequency. Higher values indicate that the participant used a higher F_0 than their own habitual F_0. Mean normalised F_0 with single Standard Deviation (SD) are given for women (red) and men (blue) according to the 11 voice protocols. Origin: own figure.

To investigate in more detail how well the study participants matched their own habitual F_0, a scatter plot indicating the individual`s performance per voice tasks with the 100% reference line was created (Figure 28). As indicated by the large spread, the third voice task was not reproducible in a considerable proportion of our study sample.

As expected in normal voices, mean voice F_0 was higher in women (Table 10). There was a significant interaction between gender and voice task ($p<0.008$, Table 11). Therefore, the effects of voice task on mean F_0 were different in women and men.

Figure 28: Normalised mean F_0 (%) in individuals per voice task



Mean fundamental frequency per voice protocol (1-11) normalised for the participant`s habitual F_0 in female (red) and male (blue) participants. The reference line at 1.00 indicates the 100% level of the overall target fundamental frequency. Higher values show that the participant used a higher F_0 than their own habitual F_0. Origin: own figure.

Table 10: Mean F_0 (Hz), SPL (dBA), jitter (%) and shimmer (%) with Standard Deviation (SD) according to voice task and protocol

Task	Protocol		mean F_0 (Hz)		mean SPL (dBA)		jitter (%)		shimmer (%)	
			women	***men***	***women***	***men***	***women***	***men***	***women***	***men***
1	Soft	Mean	229.5	127.4	61.3	64.2	1.68	1.39	7.12	6.36
		(SD)	(75.0)	(20.3)	(5.0)	(3.9)	(1.98)	(1.74)	(4.52)	(4.84)
	Medium	Mean	243.7	119.1	72.2	73.3	0.60	0.45	4.07	2.43
		(SD)	(47.7)	(11.2)	(5.7)	(3.7)	(0.55)	(0.20)	(3.59)	(1.29)
	Loud	Mean	301.8	184.5	88.5	92.0	0.39	0.25	2.71	1.65
		(SD)	(77.1)	(48.6)	(6.3)	(6.4)	(0.28)	(0.16)	(2.50)	(0.99)
2	65dB	Mean	244.9	130.5	69.9	70.3	0.74	0.66	4.64	3.00
		(SD)	(54.5)	(21.1)	(2.6)	(1.7)	(0.89)	(0.92)	(3.97)	(2.37)
	75dB	Mean	256.6	135.6	77.5	76.8	0.45	0.44	3.44	2.08
		(SD)	(61.6)	(21.4)	(2.9)	(1.5)	(0.37)	(0.66)	(3.02)	(1.78)
	85dB	Mean	275.1	152.7	85.3	84.8	0.37	0.26	2.97	1.52
		(SD)	(52.4)	(22.6)	(2.7)	(1.3)	(0.25)	(0.13)	(2.76)	(1.17)
	95dB	Mean	314.8	179.1	92.8	93.2	0.32	0.25	2.30	1.57
		(SD)	(70.5)	(31.5)	(3.8)	(1.8)	(0.23)	(0.15)	(2.18)	(1.12)
3	65dB FF	Mean	224.4	126.5	70.5	70.7	0.50	0.58	3.96	2.83
		(SD)	(35.0)	(17.0)	(3.5)	(1.9)	(0.36)	(0.57)	(2.97)	(2.52)
	75dB FF	Mean	234.1	130.0	77.2	77.2	0.46	0.32	2.83	2.04
		(SD)	(38.9)	(17.3)	(3.9)	(1.7)	(0.39)	(0.12)	(2.28)	(1.80)
	85dB FF	Mean	238.9	137.1	83.7	83.9	0.35	0.30	2.57	2.04
		(SD)	(33.2)	(19.5)	(4.0)	(2.1)	(0.28)	(0.13)	(2.43)	(1.93)
	95dB FF	Mean	261.9	159.0	89.5	90.6	0.30	0.31	2.25	2.38
		(SD)	(44.2)	(33.2)	(4.1)	(2.6)	(0.19)	(0.18)	(1.95)	(2.09)

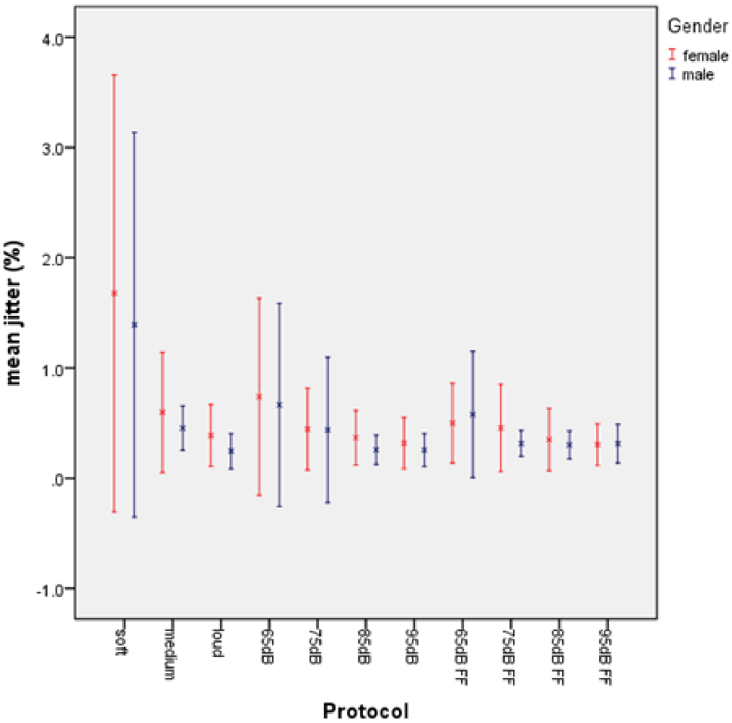
Mean F_0 (Hz), SPL (dBA), jitter (%) and shimmer (%) with single Standard deviation (SD) in women and men per voice task (1-3) and protocol (1-11). Origin: own table.

6.4.2 *Does the voice task type influence jitter and shimmer?*

Jitter and shimmer generally showed less spread at controlled voice intensity levels and also with control of voice SPL and F_0 (Table 10, Figure 29, Figure 30). Lower jitter and shimmer were always found at higher voice SPL, regardless of voice task type (Figure 29, Figure 30). However, Figure 29 and Figure 30 show that the Standard Deviation (SD) was comparable in phonations at 85 and 95dB without and with control of $F_{0.}$ Thus, in none of these protocols the measurement spread was distinctively lower.

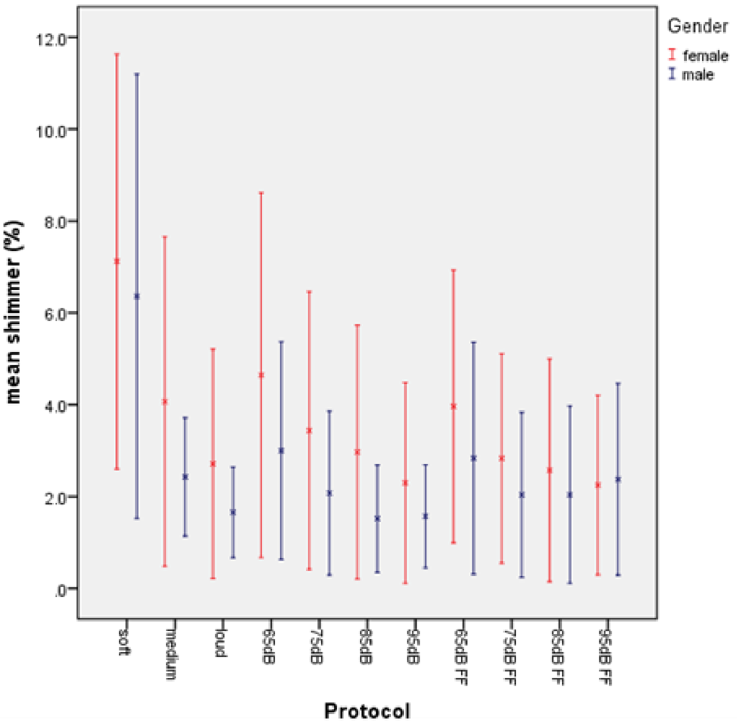
Both the voice task type (1-3) and protocol (1-11) had a highly significant influence on jitter and shimmer (p>0.001). Also, gender (p=0.002) and the participant (p>0.001) had significant effects (Table 11). For jitter and shimmer there was no interaction between gender and task (p=0.686, p=0.378). Thus, the effects of voice task were comparable in women and men for both jitter and shimmer.

Figure 29: Jitter (%) per voice task in women and men



Mean jitter (%) in women (red) and men (blue) per voice tasks. The bars indicate a single Standard Deviation (SD). From the left to the right side are shown: task 1 individual “soft”, “medium” and “loud” phonations, task 2 prescribed intensity levels of 65, 75, 85 and 95dBA (dB), and task 3 prescribed intensity levels with control of fundamental frequency (dB FF). Origin: own figure.

Figure 30: Shimmer (%) per voice task in women and men



Mean shimmer (%) in women (red) and men (blue) per voice tasks. The bars indicate a single Standard Deviation (SD). Origin: own figure.

Table 11: Influence of voice task type and protocol on voice SPL, F0, jitter and shimmer as determined by ANCOVA

Factor	Dependent variable	p-value
Task (1-3)		
	SPL (dBA)	<0.001*
	F_0 (Hz)	<0.001*
	Jitter (%)	<0.001*
	Shimmer (%)	<0.001*
Protocol (1-11)		
	SPL (dBA)	<0.001*
	F_0 (Hz)	<0.001*
	Jitter (%)	<0.001*
	Shimmer (%)	<0.001*
Gender (f/m)		
	SPL (dBA)	.004*
	F_0 (Hz)	<0.001*
	Jitter (%)	.002*
	Shimmer (%)	.002*
Participant (1-40)		
	SPL (dBA)	<0.001*
	F_0 (Hz)	<0.001*
	Jitter (%)	<0.001*
	Shimmer (%)	<0.001*
Interaction: task*gender		
	SPL (dBA)	.002*
	F_0 (Hz)	.008*
	Jitter (%)	.686
	Shimmer (%)	.378

The effects of voice task (1-3), protocol (1-11), gender (f/m) and participant (1-40) on F_0 (Hz), voice SPL (dBA), jitter (%) and shimmer (%) as determined by ANCOVA. The factor "protocol" was included as a covariable, since this depended on the task type. Effects marked with an * are considered as statistically significant. Origin: own table.

6.5 **Discussion**

The present pragmatic study shows that both the clinical task type and protocol have a statistically significant effect on voice SPL, F_0, jitter and shimmer. Thus, in clinical instrumental acoustic assessments it does matter what we ask our patients to do. To ensure basic data comparability between and even within patients a well-defined and tested standard voice protocol should be used. Ideally, this voice protocol should be doable for as many patients as possible, regardless of gender, age or training status.

6.5.1 ***How can we improve the clinical protocol in measurements of jitter and shimmer?***

As shown by Figure 27 a considerable number of our study participants were not able to control for voice SPL and keep their own F_0 at the same time. Thus, in voice clinics we can only expect unidimensional control of voice SPL from our patients, but not together with control of F_0. This also determines to which extent we can improve jitter and shimmer reliability by adjusting the recording protocol.

In the present study not all of the participants could produce and maintain 65dBA and/or 95dBA (Table 10, Figure 26). The prescribed intensity level of 85dBA without control of F_0 was matched best by both women and men. Therefore, to obtain basically comparable jitter and shimmer, a voice intensity level of 85dBA (at 10cm distance) should be used in clinical measurements. Also, as stated in chapter 5, the same vowel /a/ should always be used.

Jitter and shimmer were significantly different between women and men (Table 11), even at prescribed voice intensity levels. Therefore, as already discussed in detail in chapter 5, gender specific normative values in phonations at 85dBA with the vowel /a/ should be established.

6.5.2 *Does the voice task type influence jitter and shimmer?*

Both the voice task type and protocol had a statistically significant effect on jitter and shimmer. The measurement spread of jitter and shimmer, as indicated by the Standard Deviation, was smaller with control of voice SPL (task 2) and combined control of SPL and F_0 (task 3) as compared to soft, medium and loud voice (task 1). However, based on the Standard Deviation none of the protocols seemed to be considerably better than the others (Table 10). This might be partially explained by the small number of participants in the present study. However, this also indicates that there still might be a number of unknown sources of variance in jitter and shimmer measurements.

The present study confirms key observations from the preceding experiments described in chapters 4 and 5. In adults performing a range of phonatory tasks both jitter and shimmer are lowest at highest voice SPL, regardless of voice task, protocol or gender. This suggests that we observe a stable and presumably physiologic relation between voice SPL control and vocal fold vibration patterns. An explanatory model for this will be discussed below.

6.5.3 *Why do jitter and shimmer decrease with increasing voice SPL?*

In lower voice intensities and fundamental frequencies lower intrinsic vocal fold muscle tension and smaller vocal fold vibration amplitudes have been described (section 1.1.4; Hodge et al., 2001; Sulter et al., 1996). A lowered tension in the intrinsic vocal fold muscle might lead to a greater variability (due to increased laxness) of the covering mucosa shift. This might allow more mucosa cover movements, leading to more erratic vibration patterns and hence increased jitter and shimmer. Therefore, both parameters might indirectly indicate changes associated with laryngeal voice SPL control. Further, research has to investigate, if muscle tension or tone might be associated with jitter and shimmer. This will be focus of the research study described in chapter 8.

6.5.4 *Are jitter and shimmer perfectly useful clinical measures now?*

Based on the studies presented in chapters 4, 5 and 6, we may reasonably assume that the usefulness of clinical jitter and shimmer measurements improves by asking patients to phonate at 85dBA (10cm distance) with the vowel /a/. However, there are number of unresolved issues before a useful clinical application. First, it has to be investigated in a larger study, if these proposed measures truly improve the reliability of jitter and shimmer. This could be done by examining a larger group of healthy adults. Based on this it should be determined how many repetitions are needed for a reliable measurement of jitter and shimmer (Scherer, 1995).

However, even if the reliability of clinical jitter and shimmer measurements improves by these measures, it is unclear which voice patients are able to perform the proposed protocol. Also, the compliance might vary considerably between age groups. For example aged persons tend to use a significantly lower voice SPL during speech than younger adults, and might not be able to sustain 85dBA repeatedly (Hodge et al., 2001). Before a useful and efficient application of the proposed protocol in clinical practice, these issues have to be addressed in a larger study in voice patients and in healthy adults of different age groups.

A further main unresolved question is, that to date it is not well established what jitter and shimmer might indicate, and under which phonatory conditions (please see section 2.3). At present we do not know if clinically relevant voice properties can be measured at a voice intensity level of 85dBA. If these only appear in softer phonations, we simply would not detect the relevant information under our proposed protocol. As discussed above in section 6.5.3, a lower voice SPL might be associated with more flexibility in the mucosa cover and therefore with increased perturbation. Videolaryngoscopic examinations show a reduced flexibility of the mucosa (as indicated by a smaller or absent mucosal wave) in the presence of vocal fold pathology such as an inflammation or tumour growth (Mehta et al., 2010). Since the mucosa might be less flexible in louder phonations, pathological changes to the vibratory properties of the mucosa might not be detected under the proposed protocol. Thus, in more reliable jitter and shimmer measurements the sensitivity and validity might be lower. Therefore, in the following chapters it will be investigated what jitter and shimmer might indicate in healthy voices.

6.6 **Conclusions**

A prescribed voice SPL of 85dBA without control of F_0 was matched best by both women and men. This voice protocol should be applied in clinical measurements of jitter and shimmer to reduce the confounding effects of voice SPL and gender. Further, the vowel /a/ should always be used.

A larger study has to determine if jitter and shimmer reliability improves by these measures. Further, it has to be clarified which voice patients are able to perform this voice protocol, and if jitter and shimmer indicate useful clinical information under these measurement conditions. Also, gender specific normative values would have to be established under this protocol.

6.7 **Summary of findings**

- A prescribed intensity level of 85dBA (at 10cm distance) without control of F_0 was matched best by both women and men.

- Phonations at 85dBA with the vowel /a/ should be used in clinical measurements of jitter and shimmer to minimise voice SPL and gender effects. Future research has to clarify which patients are able to perform this, and establish normative values considering gender.

- It has to be verified, if the measurement reliability improves and if jitter and shimmer provide clinically useful information under these measurement conditions.

6.8 Summary of recommendations for clinical jitter and shimmer measurements

Clinical instrumental assessment protocol

- Sustained phonations with vowel /a/
- At a voice SPL of 85dBA (with visual feedback), at 10cm distance
- No control of F_0
- Calculation of the mean of several repetitions

Chapter 7. Do jitter and shimmer track subtle perceptual voice differences?

Instrumental acoustic voice analysis is based on the assumption that the acoustic waveform represents the vibratory characteristics of the vocal folds (chapter 2). Thus, it has been hypothesised that jitter and shimmer might be objective measures of perceptual hoarseness. However, the literature review of chapter 2 showed that the relation between instrumental perturbation and perceptual hoarseness has been described in contradictory terms.

One key methodological problem in previous studies might have been, that most of them were conducted in Type 2 and Type 3 voice signals. However, in those voice types the reliability of jitter and shimmer measurements is lowest (please see chapter 2.4). In the present chapter 7, it will be investigated if jitter and shimmer agree with perceptual voice irregularity in healthy voices (i.e. Type 1 voice signals).

7.1 Introduction

Both jitter and shimmer have been described as objective measures of the biomechanical vibratory properties of the vocal folds, which are considered central to the determination of vocal quality (Maryn et al., 2009; Mehta and Hillman, 2008; Baken and Orlikoff, 2000c). In measuring pitch and amplitude perturbation, it has been suggested that jitter and shimmer indicate different aspects of perceptual dysphonia (Martens et al., 2007; Ma and Yiu, 2006; Wolfe and Martin, 1997; Dejonckere et al., 1996; Eskenazi et al., 1990).

However, previous reports disagree on the exact nature of the relationship and have related perceptual “hoarseness” to increased shimmer (Wolfe and Martin, 1997; Dejonckere et al., 1996) and/or jitter (Eskenazi et al. 1990; Wolfe and Martin 1997; Ma and Yiu 2006; Martens, Versnel et al. 2007). For example, Dejonckere et. al. and Wolfe and Martin found a significant correlation between shimmer and “hoarseness” in pathological voices (Wolfe and Martin, 1997; Dejonckere et al., 1996). In turn, Ma and Yiu stated that jitter combined with further instrumental acoustic parameters was the best predictor of dysphonia (Ma and Yiu, 2006). Yet other authors contradict this and suggest that there is no agreement at all between instrumental acoustic and perceptual assessments. Bhuta et al. found no relation for any parameter of the GRBAS scale (Hirano, 1981) with jitter or shimmer in voice patients (Bhuta et al., 2004). Also, in healthy voices Eskenazi et. al. detected no agreement between perceptual voice ratings with jitter or shimmer (Eskenazi et al., 1990).

There are a number of reasons why the literature on this subject is so inconclusive:

(1) The types of voices that were analysed were typically dysphonic and irregular (i.e. Type 2 and Type 3 voices, please see section 2.4)

(2) The perceptual rating tool might have been inappropriate

(3) The nature of perceptual rating task was suboptimal

(1) To date, agreement between instrumental acoustic and perceptual analysis has only been reported in dysphonic voices (Martens et al., 2007; Ma and Yiu, 2006; Wolfe and Martin, 1997; Dejonckere et al., 1996). However, as discussed in chapter 2, most dysphonic voices are Type 2 and 3 acoustic signals, showing frequent changes in the acoustic wave or no apparent periodic structure. Therefore, Type 2 and 3 voice signals have been deemed as not suitable for instrumental analysis (Ma and Yiu, 2005; Carding et al., 2004; Bielamowicz et al., 1996; Titze, 1995). Minimally dysphonic or normal sounding voices, defined as Type 1 (nearly periodic or periodic) voice signals (Titze, 1995), are likely to produce most reliable jitter and shimmer (Carding et al., 2004; Bielamowicz et al., 1996; Rabinov et al., 1995).

(2) In most studies the full GRBAS scale, or some of its subparameters were used for perceptual voice analysis (Martens et al., 2007; Ma and Yiu, 2006; Bhuta et al., 2004; Wolfe and Martin, 1997; Dejonckere et al., 1996; Eskenazi et al., 1990). This results in considerable methodological problems. For example the GRBAS scale might be too coarse for comparison with instrumental perturbation measures. Jitter and shimmer indicate differences in fundamental frequency (F_0) and voice intensity (voice SPL) from one acoustic wave to the next respectively. This represents a measurement of voice irregularity at a very subtle level which might not agree with descriptions of the overall voice sound as done with the GRBAS scale. Also, there is

no evidence that the GRBAS scale parameters are the appropriate ones to relate to jitter and shimmer.

(3) The *nature* of the rating tasks has been problematic and suboptimal in previous studies. With a conventional perceptual assessment scheme such as the GRBAS scale, the examiner needs their own internal reference of "normal voice" in order to rate normal, mild, moderate or severe impairment (Carding et al., 2009; Hirano, 1981). Therefore, interrater and intrarater reliability in perceptual assessments has been criticised as limited (Kreiman et al., 1993). However, meaningful comparisons between perceptual and instrumental acoustic assessments are only possible when the reliability of both assessment types is satisfactory. Kreiman and Geratt showed that the reliability of auditory perceptual assessments are considerably improved when comparison voice samples are available to the raters (Kreiman and Geratt, 2007).

I have raised a number of critical methodological issues for a meaningful comparison between perceptual and instrumental assessments. In the experiment described in this chapter, these were addressed by a) investigating only Type 1 voice signals, and b) introducing the perceptual assessment criterion "irregularity" (analogous to instrumental acoustic irregularity), and c) using a refined perceptual assessment approach assisted by the software "Newcastle Audio Ranking" (NeAR, Appendix I, Gould et al., 2011).

For this study only vocally healthy volunteers with normal sounding voices (i.e. Type 1 voice signals) were recorded. Perceptual analysis of the voice recordings was done according to the criterion "*irregularity*" (described in detail in the methods). This parameter was chosen, since it is most likely to have a relation to an instrumental measurement of the same phenomenon. However, the perceptual concept of auditory "irregularity" is not established yet. Therefore inter- and intra-rater analysis was included in the methodology to determine if the voice experts were able to reliably assess this parameter. With the NeAR software, the examiner was presented with a set of ten voices and was asked to *rank* them from least to most "irregular" (Appendix I). Thus, the voices in a set were compared against each other and served as reference samples at the same time. This way the examiners were obliged to make a fine differentiation between completely normal sounding Type 1 voices.

7.2 **Specific study aims**

In this chapter the following study questions will be investigated in healthy voices:

(1) Do voice experts have a reliable concept of voice irregularity?
(2) Is there is a commonly understood reference for voice irregularity?
(3) Is perceptual voice irregularity related to instrumental acoustic measures of jitter (%) and shimmer (%)?

7.3 **Methods**

In the present chapter, the data described in chapter 6 was analysed further. Please refer to chapters 3 and 6 for additional details on the methods used.

In a cross-sectional single cohort study 40 vocally healthy volunteers, 20 women (mean age 28;4 years;months) and 20 men (mean age 30;1 years;months) between 20 and 40 years were investigated. All participants were recorded while phonating five seconds of /a/ under 11 voice protocols: at subjective “soft”, “normal” and “loud” voice intensity (*task 1*); at the prescribed voice intensities of 65, 75, 85 and 95dBA (with visual feedback, *task 2*); and at the prescribed intensities of 65, 75, 85 and 95dBA and with fixed fundamental frequency (F_0) (with visual and auditory feedback, *task 3*). For the present study only phonations from task 1 at “normal” voice intensity (i.e. 80 single recordings) were used.

Six voice experts ranked 10 voice recordings each in 4 gender specific subsets according to perceived “irregularity” twice in a 1 week interval. Further, rank orders according to Jitter % and Shimmer % using PRAAT were determined. Thereafter, the agreement between perceptual and objective acoustic rankings and intra- and interrater agreement were determined.

7.3.1 ***Participant inclusion criteria***

Please refer to chapter 3.

7.3.2 *Inclusion criteria for voice experts*

The voice experts in the present study were staff from the Department of Phoniatrics and Speech Pathology, University Hospital Zurich, Switzerland. Please refer to section 3.2 for detailed inclusion criteria.

7.3.3 *Perceptual voice analysis methods*

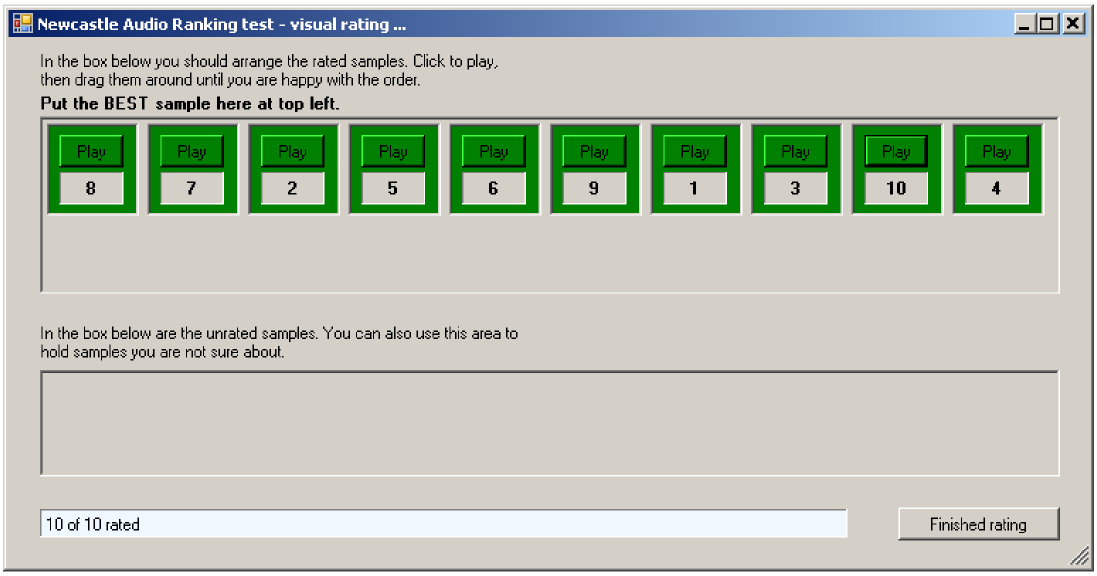
<u>The perceptual analysis software NeAR</u>

The software Newcastle Audio Ranking (NeAR, Appendix I), specifically developed by M.J. Drinnan, was used to assist perceptual analysis (Gould et al., 2011). With the software NeAR the examiner is presented with a set of voice samples and puts them in rank order, relative to the other voices (Figure 31). Voice samples in one set serve as reference recordings at the same time (Kreiman and Geratt, 2007).

Figure 31 shows a screenshot after a ranking session with 10 voices has been completed. Ranking the voices is a matter of listening to the voice samples, then arranging them in an order from least irregular (left side of the screen) to most irregular (right side). In this example voice sample 8 is ranked 1st (least irregular) and sample 4 is ranked 10th (Figure 31).

All examiners were able to listen to the voices as often as they wished. However, they were obliged to put *all* the voices in order, ties and unrated voices were not permitted. The output from the NeAR software was simply the ranking order of the voices.

Figure 31: Screenshot of NeAR software after ranking of 10 voices



Screen image of the NeAR software, after a rating session of 10 voices has been completed. In this example voice sample 8 is ranked 1st (least irregular) and sample 4 is ranked 10th (most irregular). Origin: figure by M.J. Drinnan.

<u>Expert training</u>

Perceptual voice analysis consisted of a *ranking* of 10 voices according to the criterion *"irregularity"*. All voice experts were trained together. First, the term "voice irregularity" was introduced by the examiner and defined as "irregularity or unsteadiness in the voice sound" (please see Appendix H "Oral instructions for judges"). However, irregularity was not defined with concrete terms such as creakiness, breathiness or roughness. Then, the whole group ranked 5 recordings of healthy voices according to perceptual irregularity. After that the software "Newcastle Audio Ranking" (NeAR) and its user manual were demonstrated (Appendix I). Also, every expert was able to practice with the software and to pose questions.

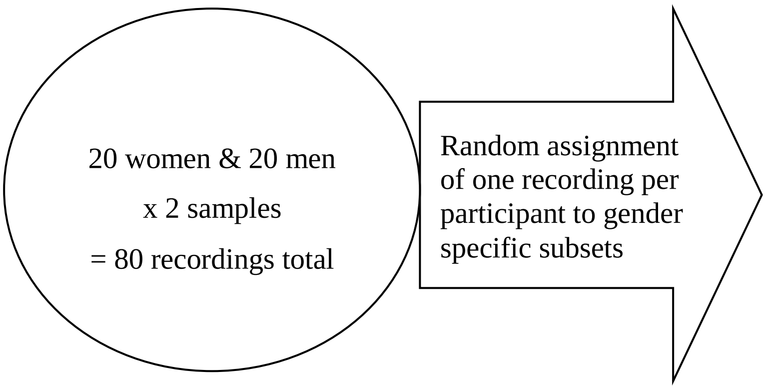
7.3.4 *Voice recording choice and preparation*

In the present chapter, the data described in chapter 6 was further analysed. Please refer to chapters 3 and 6 for further details on the methods used. Only phonations from task 1 at "normal" voice intensity were used for this study. Those were two phonations of each participant, i.e. 80 single recordings.

While theoretically any number of voices can be ranked with the software NeAR, pretests showed that 10 voices were pragmatically manageable for the examiners. Therefore, from the total pool of 80 voice recordings 4 subsets containing 10 voices each were prepared. Figure 32 shows how the recording subsets were compiled.

40 samples, one of each participant, were randomly picked and assigned to 4 subsets. Two subsets contained 10 female voices, and two subsets 10 male voices each (Figure 32). Later these subsets were used for both perceptual and instrumental acoustic analysis.

Figure 32: Method to compile subsets for the perceptual voice ranking task



Subset	Content
1	10 female voices
2	10 female voices
3	10 male voices
4	10 male voices

Method to compile 4 subsets with 10 voice recordings each. Out of 80 samples in total one recording per participant was randomly picked. These recordings were assigned to 2 gender specific subsets with 10 female or male voices each. Thereafter perceptual and instrumental acoustic analysis was conducted with the 4 subsets. Origin: own schema.

7.3.5 *Perceptual and instrumental acoustic voice assessments*

<u>Perceptual voice analysis</u>

Six speech pathologists (of the University Hospital Zurich) ranked 10 voice recordings each in 4 gender specific subsets according to perceived “irregularity”. For perceptual analysis the entire vowel phonation, as recorded before (please refer to chapter 6) was used. This task was assisted by the software NeAR (Gould et al., 2011). To allow intra-rater comparisons the identical ranking tasks were repeated after one to two weeks time.

Instrumental acoustic voice analysis

Instrumental acoustic analysis of jitter (%) and shimmer (%) was conducted with PRAAT (Boersma and Weenink 2006) using second 0.5 to 3.5 from voice onset of each phonation. Thereafter, rank orders based on the determined jitter and shimmer values were generated. The acoustic jitter (or shimmer) was treated to be equivalent to a rater in the NeAR test, i.e. the voice with the lowest jitter (or shimmer) was ranked 1, and so forth. Two voices with identical jitter or shimmer measurements are in principle possible, but the precision of the automated system makes this highly unlikely and in practice there were no ties.

7.3.6 ***Statistical analysis***

The data in this study are unusual in that for every set of rankings with 10 voices each, each score from 1 to 10 will appear just once. Thus, each set will have the same mean (5.5) and the same standard deviation (3.0). Also, all mean and SD scores will be the same for all raters. This results in two important observations: (a) there can be no systematic bias between raters, as would be possible with most rating scales, (b) many well-used coefficients of agreement, such as Pearson's correlation, intraclass correlation and weighted kappa, become equivalent under these circumstances. Therefore, the reader can place their preferred interpretation on the data, which in the interest of impartiality, was referred to as *agreement*.

Formal statistical tests can be made using Fisher's z transform, followed by the appropriate inferential test on the transformed data. The following pragmatic interpretations of agreement after Landis & Coch were used (Landis and Coch, 1977):

$\kappa < 0.4$	Poor agreement
$0.4 < \kappa < 0.6$	Moderate agreement
$0.6 < \kappa < 0.8$	Good agreement
$\kappa > 0.8$	Excellent agreement

1) *Do voice experts have a reliable concept of voice irregularity?*

Meaningful comparisons between perceptual and instrumental acoustic assessments are only possible, when the voice experts truly identify the requested voice characteristic. In this study, the concept of perceptual voice "irregularity" was newly established. Therefore, to investigate if the voice experts had a reliable concept of voice irregularity, test-retest (intrarater) agreement was measured first. A good intrarater agreement would indicate that the experts were able to apply the newly introduced perceptual assessment methods with NeAR.

2) *Is there is a commonly understood reference for voice irregularity?*

Secondly, interrater agreement was measured. A good interrater agreement would show that there was a commonly understood concept of perceptual voice irregularity. Also, this would justify comparisons between perceptual voice rankings with rankings according to jitter and shimmer.

3) *Is perceptual voice irregularity related to instrumental acoustic measures of jitter (%) and shimmer (%)?*

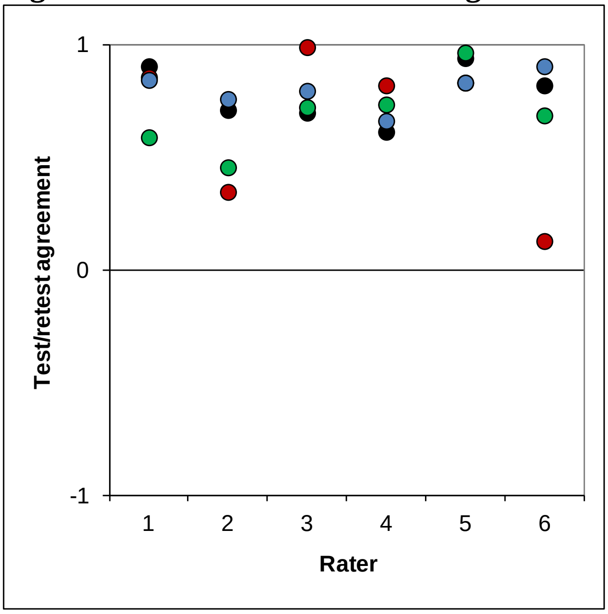
Finally, to investigate if jitter and shimmer are related to perceptual voice irregularity, the agreement between perceptual and instrumental acoustic analysis rankings was determined. A high degree of agreement would indicate that jitter and/or shimmer are numerical measures of perceptual irregularity in healthy voices.

7.4 Results

7.4.1 *Do voice experts have a reliable concept of voice irregularity?*

The overall mean (SD) for test-retest (intra-rater) agreement was 0.77 (0.11), which would be considered *good agreement* by the pragmatic scale (Landis and Coch, 1977). Figure 33 shows the test-retest agreement after 1-2 weeks separately for each rater in each of the four subsamples with ten voices. For every rater the agreement was significantly better than zero (Figure 33, $p < 0.05$). Thus, the voice experts had a stable concept of voice irregularity in healthy voices and they were able to apply the newly introduced assessment methods according to their own internal standard.

Figure 33: Test-retest intrarater agreement for 4 subsets of ten voices

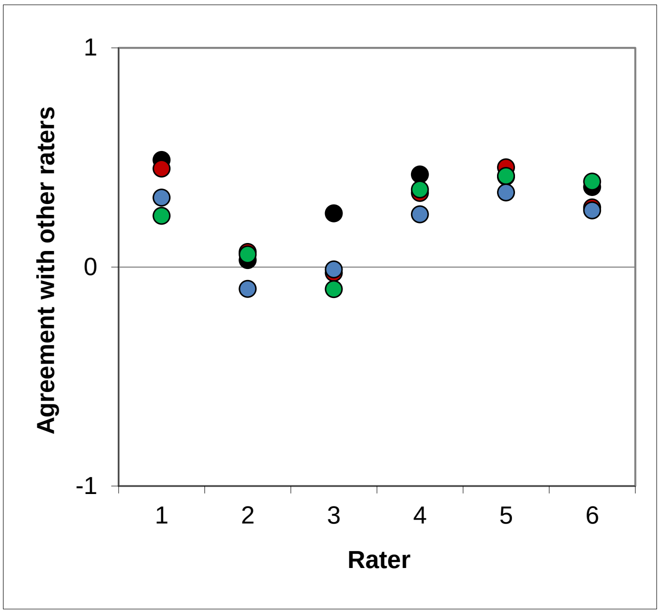


Test-retest agreement (retest after 1-2 weeks) for perceptual voice irregularity in each of the four subsamples with ten voices. Each subset is marked with a colour. Intrarater agreement is shown separately for each of the six raters. Origin: own graph.

7.4.2 *Is there a commonly understood reference for perceptual voice irregularity?*

The overall mean (SD) for interrater agreement was 0.25 (0.16), which would be considered *poor agreement* by the pragmatic scale (Landis and Coch, 1977). Figure 34 shows the agreement between six raters (interrater agreement) for four subsamples of ten voices. For each rater and subsample, his or her agreement with the other five raters was plotted. For raters 2 and 3 the agreement was not significantly better than zero (t-test, both p=0.7). Thus the concept of voice irregularity was clearly different between experts. For example raters 2 and 3, who each had good internal consistency (Figure 33), showed no agreement whatsoever with their peers (Figure 34).

Figure 34: Interrater agreement for four subsamples of ten voices

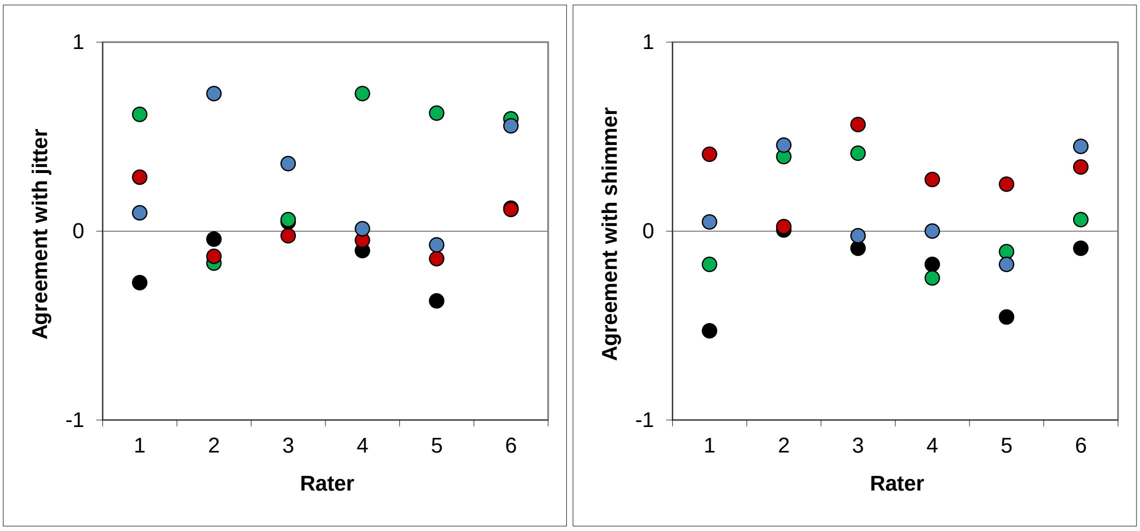


Agreement between six raters (interrater agreement) for four subsamples of ten voices. For each rater and subsample (marked in different colours), his or her mean agreement with the other five raters was plotted. Origin: own graph.

7.4.3 *Is perceptual voice irregularity related to instrumental acoustic measures of jitter (%) and shimmer (%)?*

The overall mean (SD) for agreement of perceptual irregularity with jitter was 0.18 (0.10), and with shimmer was 0.07 (0.15). These would be considered *poor agreement* by the pragmatic scale (Landis and Coch, 1977), and it would not be reasonable to use jitter or shimmer as a surrogate marker for perceptual voice irregularity. Figure 35 shows the agreement of perceptual voice irregularity with (left) jitter and (right) shimmer for four subsamples of ten voices. Each rater's agreement is shown separately. Nevertheless, considering all raters together there was a statistically significant relationship between perceived irregularity and jitter (t-test, p = 0.02). This implies that a proportion of perceptual "voice irregularity" might be related to frequency perturbations in the voice.

Figure 35: Agreement of perceptual voice irregularity with jitter and shimmer



Agreement of perceptual voice irregularity with (left) jitter and (right) shimmer for four subsamples of ten voices. Each rater's agreement is shown separately. The subsamples are marked with colours. Origin: own graph.

7.5 **Discussion**

This work was the first clinical study with the software NeAR (Gould et al., 2011). Therefore, it was not clear if all voice experts would be able to apply the new perceptual assessment protocol assisted by NeAR (Appendix I). Based on the good intrarater reliability it was assumed that all voice experts were in principle able to implement the new perceptual voice assessment procedure. The study results are discussed in detail below.

7.5.1 ***Do voice experts have a reliable concept of voice irregularity?***

As indicated by a good intrarater agreement (Figure 33), all voice experts had a stable internal reference for the newly introduced perceptual criterion “irregularity”. The experts were even able to track more subtle differences in healthy voices than usually investigated with the GRBAS scale (Hirano, 1981). Thus, perceptual analysis of discrete differences in generally normal sounding voices is in principle possible.

7.5.2 ***Is there a commonly understood reference for perceptual voice irregularity?***

The agreement between jitter and perceptual analysis for all raters together suggests that a proportion of perceptual “voice irregularity” could be related to frequency perturbations (jitter) in the voice. However, the low interrater agreement shows that the internal concept of “irregularity” remained unequal between the voice experts, even though they were instructed together (Figure 34). Thus, as already described by Bele in trained and untrained healthy voices, the acoustic phenomenon of “ normal voice” is far more complex than irregularity alone (Bele, 2005).

Kreimann et. al. suggested that the interrater reliability of perceptual assessments increases considerably, when the comparison voice samples are very similar to the target voices (Kreiman and Geratt, 2007). In this study unselected voice recordings of healthy adults were used, which might have introduced too much natural variance between the voices. A high degree of similarity between the voice samples in one set might be reached by using recordings of the same person undergoing voice changes, for example associated with voice use (Laukkanen and Kankare, 2006; Rantala et al., 2002). When presented in the NeAR test, the examiner would have to draw distinctions between the same participant`s voice at different stages of voice load. This way, the natural and confounding differences between speakers would be excluded (Kreiman and Geratt, 2007; Bele, 2005). However, it is not established yet if examiners are able to perceptually analyse such subtle differences in the same voice.

7.5.3 *Is perceptual voice irregularity related to instrumental acoustic measures of jitter (%) and shimmer (%)?*

Despite a significant statistical relationship between jitter and perceptual irregularity for the consensus of all raters, the low agreement between raters (interrater agreement) indicates that a true relationship cannot be assumed for every examiner. Thus, the correlation between jitter and the group results for perceptual irregularity appears only as a cumulative statistical effect. Based on this neither jitter nor shimmer can be considered as numerical indices of perceived irregularity in healthy voices.

One major reason for the inconsistent study results might be that the voice samples were not ideal for jitter and shimmer analysis. Even though only Type 1 voices were used, the voice recordings were not optimally controlled in terms of voice SPL. In this study the participants were instructed to phonate at individual "normal" voice intensity (voice SPL) to obtain recordings with as much natural perceptual distinctions as possible. This was done since from previous literature it was not clear if the examiners would be able at all to use "irregularity" as perceptual assessment criterion in healthy voices. However, jitter and shimmer measurements in phonations at individual "normal" loudness are highly influenced by natural differences in voice SPL between speakers (Brockmann et al., 2011). This might have masked the true relation between perceptual and instrumental assessments.

7.6 Conclusions

There is no practically meaningful agreement between perceptual and objective acoustic voice analysis in healthy voices when the usual instrumental assessment protocol is applied. However, the reliability of instrumental and perceptual assessments might be improved, which might allow a better comparison between both methods.

A more appropriate approach might be to use recordings of the same person undergoing voice changes (Kreiman and Geratt, 2007), for example associated with voice use (Laukkanen and Kankare, 2006; Rantala et al., 2002). Further, as discussed in chapter 6, jitter and shimmer reliability might be improved by asking the participants to produce a predefined voice SPL of 85dBA with the vowel /a/ (Brockmann et al., 2011). These methodological issues will be addressed in the following study described in chapter 8.

7.7 Summary of findings

- Due to limited assessment reliability the agreement between perceptual and instrumental acoustic voice examinations remains unsatisfactorily established.

- By ranking voices against each other all experts were able to indicate discrete differences in healthy voices on a finer level than usually assessed by GRBAS scale.

- A more appropriate approach to investigate the agreement between perceptual and instrumental acoustic assessments might be to use recordings of the same person undergoing voice changes, for example associated with voice use, at a predefined voice SPL of 85dBA.

Chapter 8. Do jitter and shimmer indicate subtle voice changes associated with voice use when applying an improved assessment protocol?

Until now clinical studies have not proven that jitter and shimmer are absolute or independent indices of vocal pathology or perceptual hoarseness (chapters 2 and 7). Also, in the preceding chapter 7, we were not able to verify if jitter and shimmer indicate perceptual voice differences. However, for a useful clinical interpretation of acoustic assessments, it is key to determine what jitter and shimmer might indicate. In chapters 4, 5 and 6 it was shown that voice SPL might have been a considerable confounding factor in previous studies. Further, from the literature review of chapter 2, we are aware that a useful application of jitter and shimmer might be restricted to within-participant study designs.

These issues were addressed in the present study. The main aim of the experiment presented in chapter 8 was to examine if jitter and shimmer track subtle voice changes when applying an improved instrumental voice assessment protocol (chapter 6). Healthy teachers were recorded at different times during a working day to investigate if jitter and shimmer indicate physiologic effects of voice use under these conditions. Additionally, it was assessed if jitter and shimmer agree with perceptual and subjective voice symptoms. The present study showed that jitter and shimmer might indirectly indicate changes in vocal fold tone associated with adaptation to voice use.

8.1 Introduction

Instrumental acoustic measurements have been described as most reliable in normal or near to normal sounding (i.e. Type 1) voices (Carding et al., 2004; Rabinov et al., 1995; Titze, 1995). By measuring changes from one acoustic wave to the next, it has been hypothesised that jitter and shimmer might indicate discrete vocal dysfunction in these voices (Vashani et al., 2010; Niebudek-Bogusz et al., 2007; Pribuisiene et al., 2006; Stojadinovic et al., 2002; Vieira et al., 2002). However, it is not satisfactorily established to date, if instrumental acoustic measurements reliably track discrete voice alterations (Brockmann-Bauser and Drinnan, 2011; Brockmann et al., 2011).

One main criticism raised in the present thesis is that previous research work has not adequately controlled for voice SPL in measurements of jitter and shimmer (chapters 2, 4, 5 and 6). Therefore, the effects we hope to measure with jitter and shimmer might have been confounded (Brockmann-Bauser and Drinnan, 2011). Consequently, it is not satisfactorily established in which clinical applications instrumental assessments might be most useful. Further, from the literature we are aware that meaningful jitter and shimmer measurements might be restricted to within-participant study designs (chapter 2). Therefore, in this chapter it will be evaluated if jitter and shimmer indicate discrete voice changes when applying an improved acoustic assessment protocol (chapter 6) in a clinically relevant participant group undergoing subtle voice changes.

Around half of all clinical voice patients are professional voice users, and therefore represent an important patient group (Van Houtte et al., 2011; Van Houtte et al., 2010). The most common diagnosis is "functional voice disorder", defined as vocal dysfunction without laryngeal pathology or neurogenic disease (Van Houtte et al., 2010; Altman et al., 2005). In these cases, the correct diagnosis relies on instrumental acoustic, perceptual and subjective assessment techniques (section 1.2; Altman et al., 2005; Schneider et al., 2002; Dejonckere et al., 2001).

Of professional voice users, teachers form the biggest subgroup with the highest documented risk to develop an occupational voice disorder (Bermúdez de Alvear et al., 2010; Van Houtte et al., 2010; Sliwinska-Kowalska et al., 2006). Teachers' voices have been extensively studied in clinical tests and under working conditions. A Medline literature search combining the terms "teachers" with "voice disorders" and "school" identifies a comparatively large literature base of 77 papers. Therefore, it was concluded that teachers form a highly relevant and appropriate participant group for the present study. The most significant works for the present study are summarised in Table 12.

Table 12: Summary of instrumental acoustic voice research in teachers

Reference	Studied group	Voice recording	Main Findings
Rantala and Vilkman, 1999	10 female teachers	During lessons and in breaks	Teachers with more subjective voice symptoms have a lower voice SPL, jitter and shimmer and higher F_0 than healthy teachers after a school day
Rantala et al., 2002	33 female teachers	During first and last school lesson	Higher F_0 after school day, less increase in teachers with more subjective voice symptoms
Jonsdottir et al., 2003	5 teachers	During first and last school lesson	Increase in voice SPL, the increase is less when voice amplification is used
Laukkanen and Kankare, 2006	22 male teachers	Text reading and vowel phonation before and after 6 hour working day	Increase of F_0 and voice SPL in all teachers; in teachers with more subjective symptoms more vocal fatigue, in normal teachers lower jitter after school day
Laukkanen et al., 2008	79 female teachers	Reading and vowel phonation before/ after school day	After a school day: F_0 and voice SPL are higher, jitter and shimmer are lower
Niebudek-Bogusz et al., 2007	51 female teachers with functional dysphonia	Before and after vocal loading test	Increased jitter after vocal loading
Niebudek-Bogusz et al., 2010	120 female dysphonic teachers and 30 healthy women	During standard voice assessment	Increased VHI and jitter/shimmer in dysphonic teachers
Lindstrom et al. 2010	9 female teachers 11 pupils		Increase in F_0 during day

Summary of main instrumental acoustic research in teachers during a working day and under clinical assessment conditions. Origin: own table.

Most teacher studies examined the effects of voice use in vocal loading tasks or during a working day (Table 12). The majority found an increase of fundamental frequency (F_0) and voice intensity (voice SPL) along with decreased jitter and shimmer after vocal load (Table 12).

Also, after a working day, a feeling of tiredness in the throat (Laukkanen et al., 2008), vocal fatigue, throat paresthesia and hoarseness were reported (Bermúdez de Alvear et al., 2010; Laukkanen and Kankare, 2006). When the teachers used voice amplification, the increase in voice SPL and the subjective voice symptoms were less intense; also, perceptual voice quality was rated better (Jonsdottir et al., 2003). It was concluded that instrumental acoustic measures might mirror the degree of voice load, vocal adaptation, hoarseness and subjective voice symptoms (Niebudek-Bogusz et al., 2010; Niebudek-Bogusz et al., 2008; Laukkanen and Kankare, 2006; Jonsdottir et al., 2003; Roy et al., 2002; Rantala and Vilkman, 1999). Further, it was suggested that jitter and shimmer might be an indicator of pathologic voice behaviour and of treatment efficacy in teachers (Niebudek-Bogusz et al., 2008; Rantala and Vilkman, 1999).

However, the reported results are limited for several reasons:

(1) The speech material used for instrumental analysis was not optimal
(2) Technical influencing factors might have confounded the results
(3) To date it is not clearly established, which physiologic or pathologic voice features are indicated by jitter and shimmer

(1) Most studies documented an increase of the teacher's speaking voice SPL along with a decrease of jitter and shimmer after vocal loading (Table 12). Since in all studies the instrumental acoustic measurements were conducted in phonations at "normal" voice intensity, the observed decrease in jitter and shimmer might be a side effect of the increase in speaking voice SPL (Brockmann, Storck et al. 2008; Brockmann-Bauser and Drinnan 2011). Therefore, it is not clear how jitter and shimmer respond to vocal loading, when the speaking voice SPL is adequately controlled for.

(2) One main limiting factor in measurements of jitter and shimmer is the signal-to-noise ratio (Deliyski et al., 2006; Titze, 1995). This can be optimised by controlling

the background noise, the microphone choice and the microphone-to-mouth distance (please see also section 2.2). Not all issues can be optimally addressed, for example background noise can most likely not be avoided in schools during the breaks. However, the microphone-to-mouth distance can be adjusted and should not exceed 10cm (Titze, 1995). In some studies this distance was considerably bigger, and in the most extreme case 5 feet (Lindstrom et al., 2010; Rantala and Vilkman, 1999; Schmidt et al., 1998). Also, the quality of the microphone was not always described in detail. Therefore, the reported data might have limited reliability.

(3) Previous research has not clearly established which physiologic or pathologic voice features are indicated by jitter and shimmer (chapter 2; Brockmann-Bauser and Drinnan 2011). For example, as discussed in detail in chapters 2.3 and 7, both jitter and shimmer cannot be simply considered as an index of perceptual voice quality. Therefore, comparisons with perceptual assessments or conclusions about voice function might be of limited value (Brockmann-Bauser and Drinnan 2011).

In the present study the issues raised will be addressed as follows: a) an improved instrumental acoustic assessment protocol will be applied (please see chapter 6); b) only measurements within individuals will be compared (chapter 2); c) current technical measurement guidelines will be adopted as closely as possible (chapter 2); d) to verify if teachers truly experience voice changes, perceptual voice assessments using an improved method (as discussed in detail in chapter 7) and subjective assessments will be conducted. If appropriate, the agreement between perceptual and subjective voice changes with jitter and shimmer will be determined. This might help to establish the validity of jitter and shimmer to indicate discrete voice changes.

8.2 **Specific study aims**

In this chapter the following study questions will be investigated in vocally healthy teachers:

(1) Do jitter, shimmer, voice SPL and F_0 indicate voice changes associated with voice use?

(2) Do teachers show perceptual voice changes during a working day?

(3) Do teachers experience subjective voice changes during a working day?

(4) Are perceptual and subjective voice assessments the appropriate tools to evaluate what jitter and shimmer might indicate?

8.3 **Methods**

For this study, 12 vocally healthy female teachers between 20 and 40 years (mean age 29;6 years;months) were recruited from local schools. All participants were recorded before school and after 2, 4 and 6 lessons (45 minutes duration each) of teaching. Between lessons 4 and 5 there was the main break of two hours. During this time the teachers had a voice rest.

At each point in time, the teachers answered the Voice Handicap Index (VHI) (Jacobson et al., 1996) and the question “How good is your voice” with a 100mm Visual Analogue Scale (VAS). Thereafter, every participant was recorded while phonating the vowel /a/ at “normal” voice intensity and at 85dBA. Later, the voice recordings were edited and assessed by perceptual and instrumental acoustic analysis.

8.3.1 ***Participant inclusion criteria***

Please refer to chapter 3.

8.3.2 *Voice recording protocol*

Please refer to chapter 3 for details on the equipment and calibration of the recording system and for information on the processing of all voice recordings.

All teachers performed two voice tasks and were allowed to train up to 5 minutes before the first recording. The instructions for the voice tasks were as follows:

Task 1: “Please say the vowel /a/ for 5 seconds at comfortable loudness and pitch”.

Task 2: “Please say /a/ loudly, so that the indicator showing the loudness of your voice is at the red line on the screen” (the red line indicated a voice intensity level of 85dBA). When the teachers were able to comfortably do each task, they were asked to repeat it three times while their voices were recorded. The procedure was identical at each recording point in time. Afterwards, each vowel phonation was cut out and labelled using Audacity (Mazzoni et al., 2008).

8.3.3 *Voice recording point in time*

All participants were recorded at 4 points in time: before school and after 2, 4 and 6 lessons of teaching. Each lesson had a duration of 45 minutes. Between lessons 4 and 5 there was the main break of two hours.

8.3.4 *Subjective voice assessments*

Voice Handicap Index (VHI)

At each recording session the participant’s subjective voice symptoms were investigated. For this, the Voice Handicap Index (VHI), a highly valid questionnaire to assess functional, physical and emotional aspects of vocal symptoms, was used (Carding et al., 2009; Nawka et al., 2003). The VHI has been applied in several studies to describe voice symptoms in teachers and has been shown to relate with instrumental acoustic analysis outcomes (Niebudek-Bogusz et al., 2010; Carding et al., 2009; Niebudek-Bogusz et al., 2008; Niebudek-Bogusz et al., 2007; Roy et al., 2002).

Visual analogue scale

To assess vocal wellbeing each participant was asked to answer the question “How good is your voice” by marking a Visual Analogue Scale (VAS) of 100 mm length. The left side represented “very bad” and the right side “very good” (Appendix J). Every participant did all ratings on one piece of paper to allow comparisons with earlier answers. For data analysis the degree of subjective vocal wellbeing was measured in mm; low values indicated a very good, and high values a very bad subjective judgement of vocal wellbeing (range 0 mm to 100 mm).

8.3.5 *Acoustic voice assessments*

Instrumental acoustic analysis

Instrumental acoustic analysis was conducted with the software PRAAT (Version 4.4.04). The sample used was 0.5-3.5 seconds from voice onset of each phonation (Boersma und Weenink, 2009). The following parameters were assessed:

- Jitter (%) called “Jitter (local)” in PRRAT,
- Shimmer (%) called “Shimmer (local)”,
- Fundamental frequency (Hz) called “mean pitch”,
- Voice intensity (dB) called “mean energy intensity”.

To determine calibrated voice SPL dBA the comparison method was used (chapter 3, Winholtz and Titze, 1997a).

Perceptual voice analysis

For perceptual voice analysis of all voice recordings 12 subsets with 8 samples each were prepared. Six voice experts from the University Hospital Zurich ranked the 8 recordings in each subset for perceptual “irregularity”. For this task, the entire vowel phonation, as recorded and prepared, was used.

Each subset comprised recordings of one teacher only. Included were (a) two pre-teaching recordings, (b) two recordings after 2 lessons, (c) two recordings after 4 lessons, and (d) two recordings after 6 lessons of teaching. The first and the last of the three recordings at 85dBA were always used.

The perceptual analysis methodology is identical to the approach discussed in detail in chapter 7. Again the software NeAR was used (Appendix I) and all voice experts were trained together following the same procedure as described in chapter 7 (Appendix H).

8.3.6 ***Statistical analysis***

First, mean jitter (%), shimmer (%), F_0 (Hz) and voice SPL (dBA) at normal and the prescribed voice intensity of 85dBA were plotted according to recording session for each teacher individually and for all teachers together.

For perceptual voice analysis, the mean rankings by all raters together for each teacher, and the mean rankings for all teachers together were displayed according to session. This was also done with the overall results of the Voice Handicap Index (VHI) and the Visual Analogue Scale (VAS).

Thereafter, the means (SD) for all instrumental acoustic parameters, the subjective voice symptoms and perceptual analysis at “normal” and prescribed voice intensity were determined. Analysis of Variance (ANOVA) was used to assess if the observed changes from the first to the last recording session were significant.

8.4 **Results**

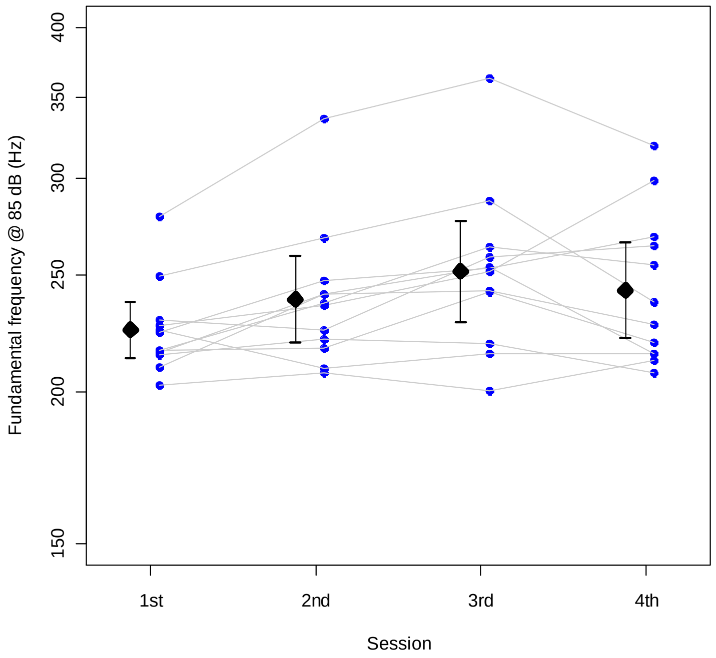
8.4.1 ***Do jitter, shimmer, voice SPL and F_0 indicate voice changes associated with voice use?***

Mean F_0 and voice SPL changes

Mean F_0 significantly (p=0.02) increased from the first to the last voice recording in phonations at normal voice intensity and at 85dBA (Table 13, Table 14, Table 15). However, mean F_0 was highest at recording session three in both loudness levels. This was before the teachers had their main break of two hours duration (Figure 36).

As expected, the teachers' mean voice SPL remained equal across sessions in phonations at the prescribed voice intensity of 85dBA. However, at normal voice intensity the speaking voice SPL significantly increased during a working day (Table 13, Table 15). As shown in Figure 37, the main increase in SPL occurred between the first and the second recording session.

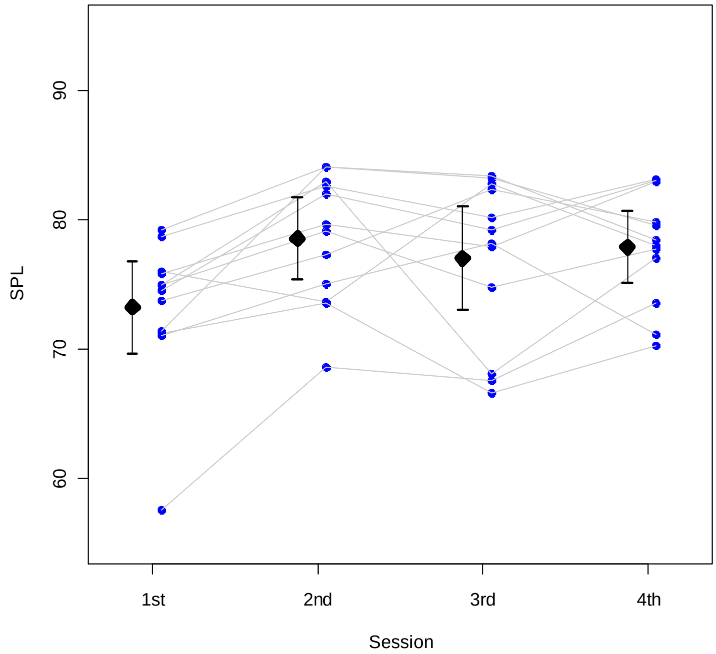
Figure 36: Increase in mean F_0 in teachers across a school day



Increase in mean fundamental frequency (F_0) in recordings at 85dBA across a school day. The blue dots show each teacher`s mean F_0 (Hz) according to session. A grey line connects the measurement results of one participant. The mean F_0 and Standard Deviation (SD) for all teachers together are indicated by the black dot with the error bar. Origin: own graph.

Figure 36 shows the increase in mean fundamental frequency (F_0) in phonations at 85dBA. Figure 37 indicates the increase in voice SPL in phonations at ”normal” voice intensity across a school day. The blue dots show each teacher`s mean F_0 (Hz) and mean voice SPL (dBA) respectively, according to recording session 1 to 4. A grey line connects the measurement results of one participant. The mean F_0, mean voice SPL and SD for all teachers together are indicated by the black dot with the error bar.

Figure 37: Change of speaking voice SPL across a school day in teachers



Change of the speaking voice SPL across recording sessions in phonations at normal voice intensity. All participants spoke significantly louder after 6 lessons of teaching than before school. However the main increase in voice SPL was between the first and the second lesson. Origin: own graph.

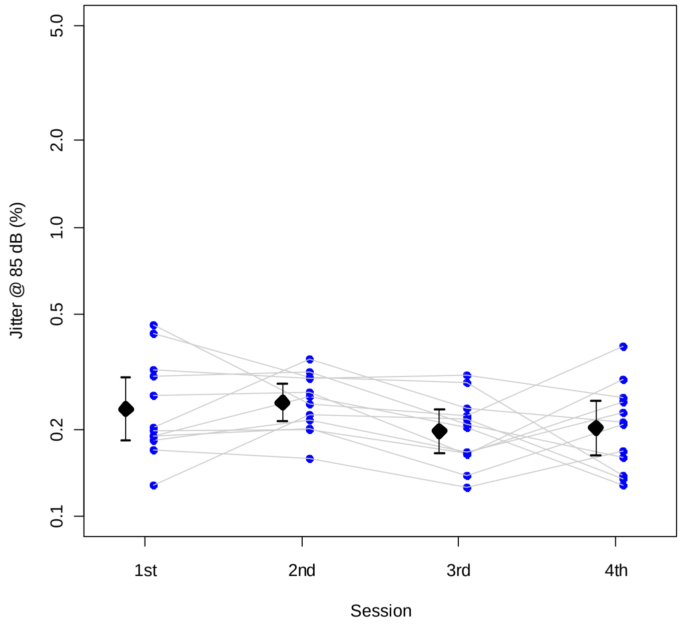
Jitter and shimmer

Jitter significantly decreased in phonations at normal intensity across a working day (Table 13, $p=0.04$). Also, at a voice SPL of 85dBA jitter decreased with time, but this result was not significant (Table 13; $p=0.1$). For shimmer, a significant decrease was observed in phonations at 85dBA ($p= 0.01$) but not at "normal" voice intensity ($p=0.09$).

One concern in the analysis of jitter and shimmer was that at normal voice intensity one teacher phonated considerably softer than her peers (Figure 37). From the previous studies we were aware, that differences in voice SPL significantly influence jitter and shimmer. Indeed, a subanalysis of the data showed, that the outlier had considerably higher jitter and shimmer than the other teachers. Therefore, the changes in the instrumental acoustic parameters were also assessed without the outlier in both phonations at normal intensity and at 85dBA (Table 15). Notably, the decrease in jitter across sessions at "normal" voice intensity was no longer significant ($p=0.16$). Also, the decrease in shimmer in phonations at 85dBA was still significant, but to a lesser extent ($p=0.04$).

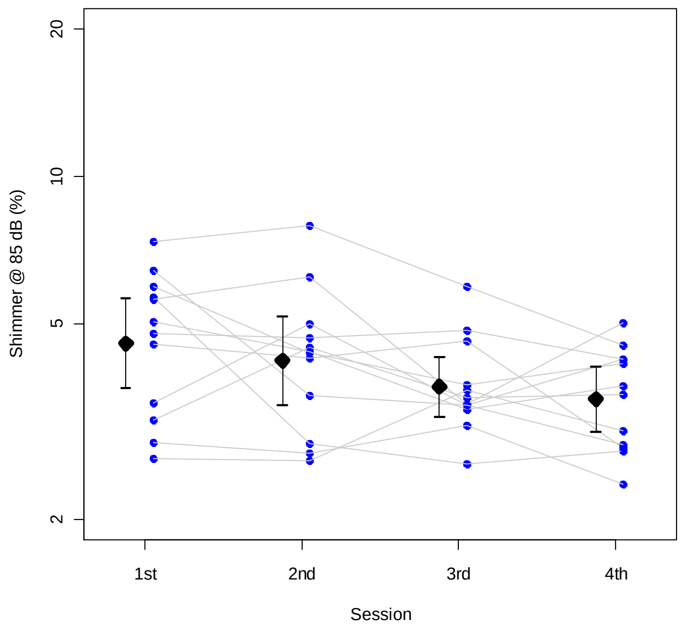
Figure 38 and Figure 39 show jitter (%) and shimmer (%) respectively according recording point in time. Table 13 displays mean values and Standard Deviations (SD) for the instrumental acoustic parameters F_0 (Hz), voice SPL (dB), jitter (%) and shimmer (%) in phonations at subjectively "normal" voice intensity from recording session 1 to 4. Analogously, Table 13 shows the mean values (SD) for all instrumental parameters in phonations at the prescribed voice SPL of 85dBA.

Figure 38: Jitter (%) in phonations at 85dBA across a school day in teachers



Jitter (%) in phonations at 85dBA across recording point in time. There was no significant change from the first to the last session. Origin: own graph.

Figure 39: Shimmer (%) in phonations at 85dBA across a school day in teachers



Shimmer (%) in phonations at 85dBA according to recording point in time. Shimmer significantly decreased across a working day. Origin: own graph.

Table 13: Mean values (SD) for F_0 (Hz), voice SPL (dBA), jitter (%) and shimmer (%) in phonations at "normal" voice intensity

		Phonations at "normal" voice intensity			
Session		**F_0 (Hz)**	**SPL (dBA)**	**jitter (%)**	**shimmer (%)**
1	Mean	203.50	73.24	0.55	7.77
	(SD)	(23.99)	(6.22)	(0.71)	(3.95)
2	Mean	220.06	78.54	0.33	6.15
	(SD)	(27.25)	(5.76)	(0.23)	(3.75)
3	Mean	234.31	77.01	0.35	6.38
	(SD)	(35.46)	(7.19)	(0.30)	(4.00)
4	Mean	230.31	77.88	0.30	5.75
	(SD)	(32.05)	(4.80)	(0.11)	(2.33)

Mean values and Standard Deviations (SD) for the instrumental acoustic parameters F_0 (Hz), voice SPL (dBA), jitter (%) and shimmer (%) in phonations at subjectively "normal" voice intensity from recording session 1 to 4. Origin: own table.

Table 14: Mean values (SD) for F_0 (Hz), voice SPL (dBA), jitter (%) and shimmer (%) in phonations at 85dBA

		Phonations at 85dBA			
Session		**F_0 (Hz)**	**SPL (dBA)**	**jitter (%)**	**shimmer (%)**
1	Mean	225.5	83.1	0.25	4.80
	(SD)	(21.1)	(2.9)	(0.13)	(1.91)
2	Mean	240.6	83.6	0.25	4.43
	(SD)	(34.3)	(2.4)	(0.07)	(1.73)
3	Mean	254.3	83.9	0.20	3.82
	(SD)	(40.4)	(3.0)	(0.07)	(1.11)
4	Mean	245.1	84.4	0.21	3.61
	(SD)	(36.4)	(2.7)	(0.09)	(1.09)

Mean values and Standard Deviations (SD) for F_0 (Hz), voice SPL (dBA), jitter (%) and shimmer (%) in phonations at a prescribed intensity level of 85dBA according to recording point in time (session 1-4). Origin: own table.

Table 15 summarises the results from ANOVA. The significance of the changes in the instrumental acoustic parameters voice SPL, F_0, jitter (%) and shimmer (%) are displayed in column two. The third column shows the results of ANOVA without a teacher who showed an exceptionally low voice SPL in the phonations at "normal" voice intensity.

Table 15: Statistical significance of the changes in voice SPL, F_0, jitter (%) and shimmer (%)

Outcome measure	p-value (all participants)	p-value (outlier excluded)
Voice SPL (85dBA)	0.21	0.3
Voice SPL ("normal" intensity)	0.004	0.016
F_0 (85dBA)	0.002	0.002
F_0 ("normal" intensity)	0.002	0.01
Jitter (85dBA)	0.10	0.11
Jitter ("normal" intensity)	0.04	0.16
Shimmer (85dBA)	0.01	0.04
Shimmer ("normal" intensity)	0.09	0.15

Significance of the changes in the instrumental acoustic parameters voice SPL (dBA), F_0 (Hz), jitter (%) and shimmer (%) with ANOVA. The third column shows the results of ANOVA without a teacher, who showed an exceptionally low voice SPL in the phonations at "normal voice intensity". Origin: own table.

8.4.2 *Do teachers show perceptual voice changes during a working day?*

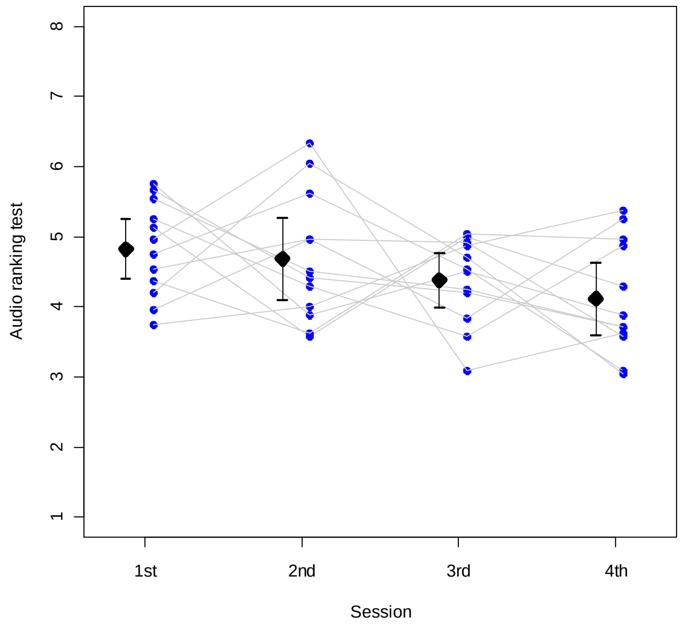
When considering the overall results of all raters and teachers together, the voice samples were ranked better the later they were recorded (Figure 40, black dots with error bar). So generally, the teachers' voices seemed to sound less irregular at the end of the working day. However, this effect was not significant (p=0.12).

Figure 40 displays the results of the voice rankings after perceptual "irregularity" by six voice experts according to recording session. Blue dots indicate the mean voice rankings of all six raters together for one teacher from best (1) to worst (8). A grey line connects the group rankings for one teacher according to recording point in time (1 to

4). The black dots with the error bars represent the mean voice rankings for all teachers together.

As indicated by the blue dots in Figure 40 there was a considerable spread in the group voice rankings for each teacher. Some teachers' voices sounded even more irregular with time or showed maximum irregularity at the second recording session (grey line, Figure 40). These unequal perceptual acoustic results during a day between teachers might indicate different clinically relevant reactions to voice load. Thus, in the averaged group results of all teachers, relevant differences *between* individual voices might not appear.

Figure 40: Voice rankings after perceptual irregularity by six voice experts according to recording point in time



Voice rankings after perceptual "irregularity" by six voice experts according to recording point in time (1 to 4). Blue dots indicate the mean voice rankings (of all 6 raters together) for one teacher from best (1) to worst (8). A grey line connects the group rankings for one teacher according to recording session. The black dots with the error bars represent the mean voice rankings with Standard Deviation (SD) for all teachers together. Origin: own graph.

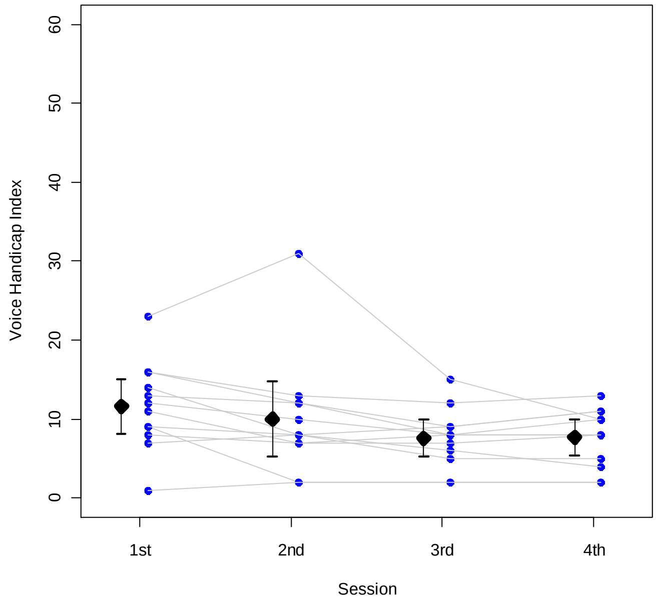
8.4.3 *Do teachers experience subjective voice changes during a working day?*

The overall Voice Handicap Index (VHI) score decreased from a mean of 11.5 (SD 5.8) in the first recording session to a mean of 7.1 (SD 3.5) in the last recording session (Figure 41). This change was highly significant ($p=0.007$). Similar results were found for the question “How good is your voice”. The mean VAS scale score for all teachers together improved from a mean of 56.6 mm (SD 17.2) to a mean of 63.9 mm (SD 16.8) in the last session (Figure 42). However, this result was not significant ($p=0.41$).

This might give the impression that the subjective voice symptoms decreased across recording sessions in all participants. However, as shown in Figure 41 and Figure 42, the results varied considerably between the teachers. Some even found their voice to worsen with teaching, or had the best subjective voice function at the 2^{nd} or 3^{rd} recording session. Especially Figure 42 shows how different the teachers estimated their vocal well-being across sessions. Therefore, as observed in the perceptual voice analysis, in the group results for subjective voice assessments clinically relevant differences between individuals might disappear.

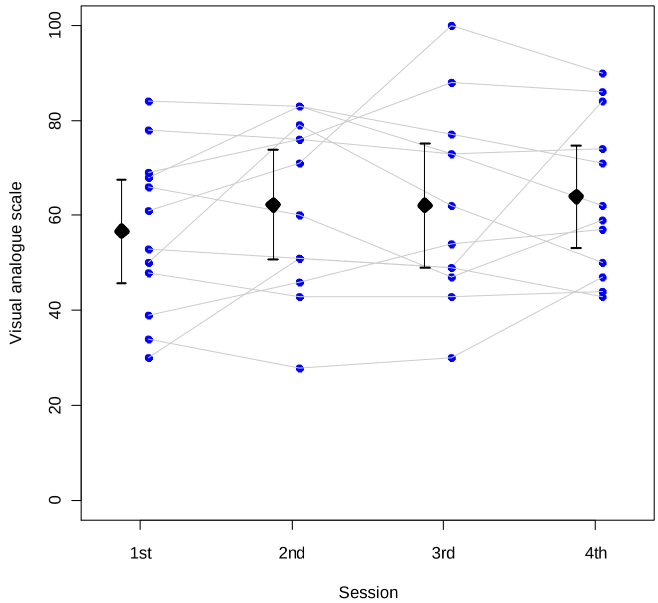
Figure 41 and Figure 42 show the subjective voice symptoms of the teachers as measured by the Voice Handicap Index (VHI) and the Visual Analogue Scale (Figure 42) according to recording session. Each teacher’s results are shown by blue dots and connected by the grey line. The black dots with the error bar indicate the overall results for all teachers together according to recording session (1 to 4).

Figure 41: Subjective symptoms in teachers during a working day measured by Voice Handicap Index



Subjective voice symptoms of the teachers as measured by Voice Handicap Index (VHI) according to recording session (1 to 4). Each teachers` results are shown by blue dots and connected by the grey line. The overall VHI results for all teachers together (black dots with the error bar) significantly improved from the first to the last session. Origin: own graph.

Figure 42: Subjective voice symptoms in teachers measured by visual analogue scale



Subjective voice symptoms as assessed by the question "How good is your voice" and a 100mm Visual Analogue Scale (VAS) scale. The blue dots indicate the mean VAS results of one teacher according to recording session (1 to 4). Higher values (indicated in mm) represent better subjective ratings. Grey lines connect the results for one teacher. Black dots with error bars indicate the overall results for all teachers together. Origin: own graph.

8.4.4 *Are perceptual and subjective voice assessments the appropriate tools to evaluate what jitter and shimmer might indicate?*

As shown by Figure 40 to Figure 42 the perceptual and the subjective assessment results show a considerable spread. Thus, in the group results of all teachers together distinct and probably clinically relevant differences between the participants might be averaged. However, as discussed in detail in chapter 7, the perceptual analysis approach might also not be adequate to investigate healthy voices. Also, the phenomenon of subjective voice wellbeing might be far more complex than the question used in this work.
Based on this the perceptual and subjective assessment results for all teachers/ raters together appear not suitable to investigate what jitter and shimmer might measure. Therefore, further comparisons between perceptual and subjective assessments with instrumental analysis were not conducted.

8.5 Discussion

8.5.1 *Do jitter, shimmer, voice SPL and F_0 indicate voice changes associated with voice use?*

In the present study refined voice assessment procedures were applied to determine if jitter and shimmer measure discrete changes associated with voice use in vocally healthy teachers. Jitter and shimmer decreased across a working day in phonations at uncontrolled and controlled voice intensity. However, this result was only significant for jitter or shimmer respectively. One reason for this might be, that in this study only a small sample of twelve teachers was examined. Further, as shown by the assessment of the outcomes with and without one teacher (outlier) who phonated with an unusually soft voice, voice SPL seems to remain a considerable confounding factor in instrumental acoustic measurements. Therefore, an investigation in a greater sample of teachers might show clearer results regarding both jitter and shimmer. Another difficulty in interpreting the presented study results is that we do not know the typical variation in teacher`s voices during a day without vocal loading.

However, a comparison to the existing literature allows preliminary conclusions. As already reported in a number of previous works (Table 12), the present study also found an increase in F_0 and voice SPL at normal voice intensity was found in teachers across a working day. Further, as expected, voice SPL remained stable in phonations at the prescribed level of 85dBA, but again F_0 significantly increased across recording sessions. Based on this we conclude that in our sample the teachers responded to voice load across a working day, and that jitter and shimmer might have tracked these changes associated with voice use.

The decrease of jitter and shimmer with increased F_0 and voice SPL has been interpreted as physiologic adaptation to vocal load or “vocal warm up” (Laukkanen et al., 2008; Laukkanen and Kankare, 2006; Rantala et al., 2002). In the present study in vocally healthy teachers a general improvement of subjective voice symptoms and of perceptual voice irregularity with the quantity of teaching was also found. This clearly supports the hypothesis of a physiologic adaptation to voice load. Also, the finding of highest F_0 at session 3 before the main break supports this concept. Thus, jitter and shimmer might indirectly indicate voice adaptation; the potential physiologic mechanisms underlying this will be discussed in detail below.

8.5.2 *Do teachers show perceptual voice changes during a working day?*

When considering the overall results for the consensus of all raters and teachers together, the voice samples were ranked better the later they were recorded. However, this result was not significant. One reason for this might be that in this study only twelve teachers were examined. However, as already found in chapter 7, the voice experts again varied considerably in their judgement of voice “irregularity” (Figure 40). This was the case despite the use of refined assessment methods and more suitable voice recordings for perceptual analysis (please see chapter 7). Thus, interrater comparability remains a main limiting factor in perceptual assessments of normal voices.

8.5.3 *Do teachers experience subjective voice changes during a working day?*

In this study, all teachers had a significant improvement of subjective voice symptoms in the Voice Handicap Index (VHI). Also, with the question “How good is your voice” an improvement of vocal well-being was found, but this was not significant. This seems to support the hypothesis, that vocally healthy teachers show a successful voice adaptation during a working day (Laukkanen et al., 2008; Laukkanen and Kankare, 2006; Rantala et al., 2002). However, as already discussed for the perceptual voice rankings, a considerable spread was found in the teachers' answers. For example, a number of participants found their voices to worsen with time (Figure 42). Thus, similar to instrumental acoustic voice analysis, comparisons of subjective voice symptoms between individuals might be of limited value.

8.5.4 ***Are perceptual and subjective voice assessments the appropriate tools to evaluate what jitter and shimmer might indicate?***

In a study by Niebudek-Bogusz et.al., it was reported that jitter and shimmer might indicate voice quality and vocal symptoms (Niebudek-Bogusz et al., 2010). However, in the present work a considerable spread was observed in the perceptual and subjective voice assessment results for all teachers/ raters together (Figure 40 to Figure 42). As discussed in detail in chapter 7, normal voices might be far more complex than "irregularity" alone. Thus, our perceptual analysis approach in this study might have been too one-dimensional to investigate even discrete differences in healthy voices. The same issue could be true for the subjective voice wellbeing, which might be a far more complex phenomenon than a rating of how "good" a voice is.

However, even if the perceptual or subjective assessment approach would have been perfectly adequate, there still remains a considerable methodological problem: for the group comparisons, distinct differences between teachers' voices were averaged; these differences might be clinically relevant. Based on this, the perceptual and subjective assessment methods of the present study appear not to be a suitable benchmark to investigate what jitter and shimmer might measure in teachers during a working day. Based on the present data it cannot be assumed for all teachers, that lower jitter and shimmer indicate improved perceptual voice quality or subjective vocal symptoms.

8.5.5 ***What might jitter and shimmer indicate?***

Instrumental acoustic measurements still have limited reliability and validity. Nevertheless, the present results and a comparison to the existing literature provide preliminary evidence that jitter and shimmer might be associated with muscle tone in the vocal folds. Videolaryngoscopic and aerodynamic examinations in healthy adults show that an increase in F_0 and voice SPL are associated with an increased tone in the vocal folds (Hodge et al., 2001; Sulter et al., 1996). This might lead to a stiffer and more stabilised vocal fold, and hence reduced irregularity in the vibration patterns (Laukkanen et al., 2008).

Studies in dysphonic teachers seem to support this hypothesis. In these, F_0 increased less as compared to healthy teachers, and both jitter and shimmer rose over a working day (Niebudek-Bogusz et al., 2008; Niebudek-Bogusz et al., 2007; Laukkanen and Kankare, 2006; Rantala et al., 2002). Thus, jitter and shimmer might indirectly indicate discrete voice alterations associated with muscle tone changes. We discussed a similar model to explain the dramatic increase in jitter and shimmer with decreasing voice SPL (chapter 6; Brockmann et al., 2008). Therefore, instrumental measurements might be a relevant assessment tool to discover discrete changes in muscle tone. In voice clinics this might help to identify patients suffering of functional voice disorders or pathologic fatigue. However, before further conclusions are drawn, the proposed hypothesis should be verified in voice patients with impaired muscle tone regulation.

8.6 **Conclusions**

When applying an improved assessment protocol, jitter and shimmer track voice changes associated with vocal load in healthy teachers. Both jitter and shimmer decrease, subjective voice symptoms improve and perceptual voice irregularity has a tendency to decrease during a working day. These observed voice changes might be explained by physiologic adaptation to vocal load. As indicated by the rise in F_0 and voice SPL during a working day this might be associated with an increase in muscle tone in the vocal folds. Therefore, decreased jitter and shimmer could be related to a higher muscle tone and physiologic vocal adaptation.

Both perceptual and subjective assessment results show a considerable spread. This hinders meaningful comparisons with instrumental acoustic measures. Based on the present study results, jitter and shimmer cannot be considered as objective measures of subjective and perceptual voice symptoms.

8.7 Summary of findings

- Jitter and shimmer track voice changes associated with vocal load when improved instrumental assessment methods are used.

- Both jitter and shimmer decreased whereas F_0 and voice SPL increased, subjective voice symptoms improved and perceptual voice irregularity reduced. This might be explained with physiologic vocal adaptation, associated with an increase in muscle tone of the vocal folds.

- Since both perceptual and subjective assessment results show a considerable spread between teachers/ examiners respectively, the association with jitter and shimmer remains unclear.

Chapter 9. Clinical and theoretical implications of this work

The ideal clinical voice assessment tool is readily available and reliable under all circumstances, be it in a (noisy) office or at the patients' bedside. Further, it is easy to interpret and gives objective information. While it was hoped that instrumental acoustic voice assessments might fulfil these criteria (Mehta and Hillman, 2008; Titze, 1995), recent research identified a number of methodological problems in clinical measurements of jitter and shimmer (Brockmann-Bauser and Drinnan, 2011). Instrumental acoustic measurements are still significantly confounded by factors associated with the measurement technique, such as the voice recording and analysis equipment or background noise (Maryn et al., 2009; Deliyski et al., 2006). Also influencing factors related to the usual voice recording procedure have not been sufficiently characterised for an efficient control in voice clinics. Patients are usually instructed to say the vowel /a/, /o/, or /i/ at "comfortable loudness and pitch". However, the present work showed that the participants` speaking voice SPL, F_0 and gender or even the vowel choice might confound both jitter and shimmer (Brockmann-Bauser and Drinnan, 2011). Also, our understanding of the physiologic and pathologic voice features, which might be indicated by jitter and shimmer, is too limited for a valid clinical interpretation.

This research work was designed to investigate the confounding effects of voice SPL, F_0, gender and vowel and how these might be reduced in clinical measurements of jitter and shimmer. Further, the present research was developed to contribute to our knowledge base of what jitter and shimmer might indicate. Based on this, useful applications of jitter and shimmer are evaluated. Below, each study aim will be considered in turn.

9.1 How important are the effects of voice SPL, F_0, gender and vowels in clinical measurements of jitter and shimmer?

In a typical clinical voice task, all factors voice SPL, F_0, gender and vowel have a significant effect on jitter or shimmer. It was shown for the first time that current clinical guidelines are not sufficient to control for these effects in clinical practice (chapter 4 and 5).

Differences in voice SPL between individuals had by far the greatest impact. The effect was so strong, that clinical jitter and shimmer measurements appear not to be meaningful without adequate control of voice SPL. By comparison, vowel and gender effects were considerably smaller, but still clinically important. Surprisingly, F_0 had a relatively small statistical effect. However, in the data presented F_0 was systematically higher in women and naturally increased with voice SPL in both genders. Since changes in F_0 could not be fully controlled for in all the studies presented here, F_0 differences might still contribute to the observed gender and voice SPL effects.

9.2 How can we improve clinical measurements of jitter and shimmer?

A clinical voice task should be doable for most patients and give results as reliable as possible. In chapter 5 it was shown that all factors voice SPL, F_0 and vowel have a significant impact on jitter and shimmer in voice clinics. Vowel effects may be controlled for by always using the same vowel. The vowel /a/ was proposed, since it is easy to reproduce and control. However, due to language and accent differences, the articulation of the vowel /a/ may vary along the front-back phonetic dimension. To date we do not know if this affects jitter or shimmer. Therefore, clinicians should ask the patients to produce /a/ forward, and check whether they are able to do this.

In chapter 6, it was investigated which voice protocol might be used in clinical measurements of jitter and shimmer. A range of voice tasks was tested in untrained vocally healthy women and men. The best performance was observed in phonations at the predefined voice intensity level of 85dBA, without control of F_0. Whilst the increase in F_0 is a physiologic phenomenon associated with the rise in voice SPL, this restricts the phonatory tasks we may ask from our patients to unidimensional control of voice

SPL. Also, a considerable proportion of participants could not phonate at 65dBA or 95dBA. Therefore, patients should phonate at a predefined level of 85dBA in clinical measurements of jitter and shimmer.

Based on the presented studies it can be assumed that the reliability and probably also the sensitivity of clinical jitter and shimmer measurements might be considerably improved by always using a prescribed intensity level of 85dBA with the vowel /a/. As discussed in chapter 5, for shimmer the sensitivity could double, i.e. changes two times smaller might be recognised (Brockmann et al., 2011). However, this is still hypothetical and has to be confirmed in a larger clinical study. Based on this, it might be determined how many repetitions should be done for a reliable clinical measurement of jitter and shimmer.

Summary of clinical instrumental assessment protocol

- Sustained phonation with vowel /a/
- At a voice SPL of 85dBA (with visual feedback), at 10cm distance
- No control of F_0
- Calculation of the mean of several repetitions

9.3 What can clinical jitter and shimmer measurements tell us about the human voice?

One key assumption about instrumental acoustic measurements is that these might objectively indicate perceptual hoarseness (Mehta and Hillman, 2008). However, the literature review of chapter 2 showed that previous studies reported contradictory evidence on the relation between perceptual and instrumental voice analysis. In the present work this question was investigated by two separate experiments in vocally healthy adults. In both studies, refined perceptual assessment methods assisted by the software Newcastle Audio Ranking (NeAR) were used (Gould et al., 2011). Additionally, in the second experiment (chapter 8) the improved instrumental acoustic measurement procedure recommended here was applied (chapter 6). Despite the use of refined assessment methods there was no practically meaningful agreement between perceptual voice irregularity with jitter or shimmer in both studies.

One methodological problem was that the voice experts varied considerably in their judgement of perceptual voice "irregularity". As already stated by Kreimann et. al., interrater reliability remains a main limiting factor in perceptual assessments (Kreiman and Geratt, 2007). Therefore, despite a correlation between perceptual irregularity and jitter for the consensus of all raters together, comparability between perceptual and instrumental irregularity cannot be automatically assumed in all cases. Based on this, we have to conclude that jitter and shimmer do not objectively measure perceptual voice irregularity. This is in agreement with our previous report in dysphonic voices (summarised in section 2.3) and highlights that the acoustic phenomenon of "voice" is far more complex than regularity alone.

Nevertheless, a comparison of the overall results to the existing literature base provides preliminary evidence that both jitter and shimmer might be associated with muscle tone in the vocal folds. In all studies described in chapter 4, 5, 6 and 8, jitter and shimmer were always lower in higher voice intensities, regardless of voice task. Videolaryngoscopic and aerodynamic examinations in healthy adults show that a rise in voice SPL and F_0 are associated with an increased tone in the vocal folds (Hodge et al., 2001; Sulter et al., 1996). This might lead to a stiffer and more stabilised vocal fold, and hence reduced irregularity in the mucosal vibration patterns (Laukkanen et al., 2008).

Thus, reduced jitter and shimmer might indirectly indicate an increase in vocal fold tone associated with increased voice SPL.

This hypothesis is supported by our observations in vocally healthy teachers. In this group increased voice SPL and F_0 after a working day were associated with lower jitter and shimmer. Also, overall subjective voice symptoms improved and perceptual voice irregularity tended to decrease. Laukkanen described these voice changes as physiologic adaptation to vocal load (Laukkanen and Kankare, 2006; Laukkanen et al., 2008). As indicated by increased F_0 and voice SPL in the examined teachers, the observed voice adaptation might also correspond with an increase in muscle tone. Again, decreased jitter and shimmer would be related to increased muscle tone, this time associated with physiologic vocal adaptation. Findings in dysphonic teachers seem to confirm the presented hypothesis: in dysphonic teachers F_0 increases less as compared to healthy teachers, and both jitter and shimmer tend to rise over a working day (Niebudek-Bogusz et al., 2008; Niebudek-Bogusz et al., 2007; Laukkanen and Kankare, 2006; Rantala et al., 2002).

9.4 What are useful applications for jitter and shimmer measurements?

Currently, both jitter and shimmer are widely used in clinical voice assessments to assist the diagnosis of voice disorders, and to document and evaluate intervention success (Brockmann-Bauser and Drinnan, 2011). However, despite the presented refinement of the measurement protocol (chapter 6) and improvements to the measurement technique (Boersma, 2009; Deliyski et al., 2006), jitter and shimmer still clearly have limited reliability (Brockmann-Bauser and Drinnan, 2011). For example in the presented study of chapter 5 there was a considerable variation in clinical measurements of jitter and shimmer that could not be explained by the factors voice SPL, F_0, gender and vowel. This highlights that we are far from understanding all factors influencing jitter and shimmer.

This has considerable implications for the clinical use of jitter and shimmer, and derived measures such as the Dysphonia Severity Index (Wuyts et al., 2000). Before key questions regarding measurement reliability and further influencing factors have not been better answered, jitter and shimmer should not be used to rate intervention success.

Also, a useful application might remain restricted to within patient measurements to track voice changes. Given the limitations to date, jitter and shimmer also appear not as valid tool to supplement an initial voice diagnosis.

Clinical measurements of jitter and shimmer may be significantly improved when patients phonate at a prescribed voice SPL of 85dBA and always use the vowel /a/. The present study in teachers showed that jitter and shimmer might indicate subtle voice changes associated with voice use under these measurements conditions (chapter 8). Further, lower jitter and shimmer might be associated with increased vocal fold tone. This would be highly relevant clinical information for the diagnosis in normal sounding patients without a visible voice pathology (Schneider et al., 2002). This is often the case in so-called functional voice disorders which are present in around 30% of the clinical caseload (Altman et al., 2005; Van Houtte et al., 2010; Van Lierde et al., 2010). Also, early neurologic voice disorders are characterised by subtle changes to vocal fold tone regulation which are difficult to detect by other diagnostic means (Schneider et al., 2002). Refined measurements of jitter and shimmer might indicate subtle changes in vocal fold tone and thereby close a diagnostic gap (section 1.2). However, a number of essential questions have to be addressed in future research. These will be discussed in the following section.

9.5 **Future research directions**

As discussed in chapter 2, improving the reliability of jitter and shimmer is key to establishing their validity (Brockmann-Bauser and Drinnan, 2011). Results of the present study suggest that clinical measurements might be more meaningful, when the patients phonate at 85dBA and always use the vowel /a/. However, before a useful and efficient clinical application of jitter and shimmer, a number of basic questions have to be addressed.

The first obvious research direction would be to investigate in a larger clinical study, if the reliability of jitter and shimmer measurements truly increases when applying the proposed improved instrumental assessment protocol. Based on this, it has to be determined how many repetitions should be made for obtaining reliable clinical jitter and shimmer measurements (Scherer, 1995). Also, it should be assessed which voice patients are able to perform this voice task. Since gender effects were still present in phonations with control of voice SPL, also gender specific normative values should be investigated under the proposed protocol.

Our knowledge base regarding further physiologic factors underlying jitter and shimmer remains limited. For example, it was not possible to fully explore the effects of F_0, since our study participants were not able to control for F_0 and voice SPL at the same time. Also, we may expect differences in jitter and shimmer related to age groups or voice training status (Stathopoulos et al., 2011). These issues have to be addressed in future works to understand what jitter and shimmer might indicate and when they are pathologic. Given that, an essential research direction for a useful clinical application would be to determine pathology thresholds under the new assessment protocol.

A main unresolved clinical problem remains: to date it is not sufficiently described, what jitter and shimmer might indicate, and under which phonatory conditions. The findings presented show, that differences between healthy speakers can be reduced considerably if we use phonations at a predefined voice SPL of 85dBA. However, this does not necessarily mean, that this voice intensity level is the best for detecting discrete voice alterations or early pathology. As discussed in section 6.5, pathologic changes in the vibratory properties of the vocal folds might only be detectable in softer phonations. Given this, an increased reliability for jitter and shimmer in phonations at 85dBA might come at the expense of measurement sensitivity and validity. This might be addressed by investigating jitter and shimmer in patients with discrete pathology in the vocal fold mucosa such as in severe laryngopharyngeal reflux. Comparisons before and after successful medical treatment, in soft phonations and at a prescribed level of 85dBA, might give insight if and at which voice intensity level jitter and shimmer might indicate mucosal pathology.

Further, the proposed hypothesis that jitter and shimmer indirectly indicate muscle tone changes has to be verified in patients. This might be done by comparing jitter and shimmer before and after vocal loading tasks, or by assessing patients with neurologic voice disorders such as dysarthrophonia or myasthenia gravis.

As shown by the comparisons between perceptual and instrumental analysis the human voice is a far more complex phenomenon than irregularity alone. Thus, more elaborate models to process the acoustic data such as nonlinear dynamics or chaos analysis might be more successful in describing normal and pathological voices (Shao et al., 2010; Mehta and Hillman, 2008). However, these methods also have to be further investigated for a useful routine clinical application (Mehta and Hillman, 2008). All in all, it is clear that instrumental acoustic measurements of jitter and shimmer have to be developed considerably before a truly meaningful routine clinical application.

A further set of unanswered questions is associated with the use of steady state vowels in acoustic voice assessments. Even though this was recommended to reduce confounding effects associated with connected speech, it is still unclear if vowel phonations are a sufficiently representative model for the habitual voice function of a patient. Speech related characteristics such as hard glottal attacks considerably contribute to the perceptual phenomenon of "dysphonia". These may occur more often or stronger in habitual speech than in isolated vowel phonation. This leads us to another main methodological question: should instrumental acoustic assessments investigate the most optimal phonation (such as vowels), or phonations that represent habitual voice function best? Clinical measurement protocols may even depend on the purpose of the assessment. This calls for further studies into what jitter and shimmer might indicate, and under which measurement conditions.

9.6 Numbers are numb- a personal note why we might love them so much

In the course of this research work I was surprised to find out how little evidence underlies the broad clinical application of jitter and shimmer. Both measures have been internationally used in almost every area of voice diagnostics, documentation and research (Brockmann-Bauser and Drinnan, 2011).

It seems that clinicians (including me before I started the present project) try to base decisions on a measure that is still unreliable and has no clear association to voice pathology. Considering this, jitter and shimmer are numbers that seem to tell us little about the human voice to date. So why are jitter and shimmer so attractive? Porter described that a society uses numbers to remove subjectivity from observations (Porter, 1995). However, as already concluded by Rabinov in 1995, objectivity is not a good reason to prefer instrumental measures over perceptual voice analysis, an assessment method with inherent subjectivity through the examiner (Rabinov et al., 1995).

Can a look at the daily clinical routine probably offer answers? Voice function is highly different between individuals, resulting in a considerable spread in literally every voice measure that is taken: be it the perceptual voice impression (Bele, 2005; Dedivitis et al., 2004), the vocal fold closure or mucosal wave (Sulter et al., 1996). Our understanding of normal voice production is still limited, so a differentiation between normal and pathologic often is difficult. The clinician is faced with patients who suffer enormously from voice problems, but there might not be a clear hypothesis about the problem (Kleemola et al., 2011; Jones et al., 2006). Thus, the diagnosis and the treatment approach is at the discretion of the clinician, and there are decisions to be made. In my view, in this situation numbers are tempting, providing seemingly objective information about the highly complex phenomenon of the human voice. Additionally, this fulfils the requirements to document treatment outcomes and to provide expert explanations.

There is no simple way to address the issues raised in this work. We may possible be on the wrong course using jitter and shimmer, and more complex approaches such as chaos analysis might provide more useful information (Mehta and Hillman, 2008). Or refined perceptual assessments might possibly tell us more about the human voice than any instrumental measure (Rabinov et al., 1995). Notably, with assistance of the software NeAR, in the present study all voice experts were able to indicate discrete differences in healthy voices on a finer level than usually assessed by GRBAS scale (Hirano, 1981; Gould et al., 2011).

Meanwhile, from a daily clinical perspective it might come down to the question: what is vocal pathology, when we can`t see or hear anything wrong? After several years of research and clinical experience with voice patients I would ultimately say: voice pathology is what the patients tell us is wrong with *their* voice. Then the clinician should use robust and meaningful diagnostic tools, to find out *what* might help the patient. And that makes it all so fascinating.

Bibliography

Akerlund, L., Kitzing, P., Petersson, G. and Linder, H. (2009) *Dysphonia*. (retrieved September 11, 2010, from http://www.dysphonia.certec.lth.se/)

Altman, K., Atkinson, C. and Lazarus, C. (2005) 'Current and emerging concepts in muscle tension dysphonia: a 30-month review', *Journal of Voice,* 19, (2), pp. 261-7.

Arjmandi, M., Pooyan, M., Mikaili, M., Vali, M. and Moqarehzadeh, A. (2011) 'Identification of Voice Disorders Using Long-Time Features and Support Vector Machine With Different Feature Reduction Methods.', *Journal of Voice,* 25, (6), pp. e275-89.

Baken, R. (1992) 'Electroglottography', *Journal of Voice,* 6, (2), pp. 98-110.

Baken, R. and Orlikoff, R. F. (2000a) 'Amplitude Perturbation (Shimmer)', in *Clinical Measurements of Speech and Voice.* 2nd ed Albany, New York: Thomson Delmar Learning, pp. 130-137.

Baken, R. and Orlikoff, R. F. (2000b) 'Frequency Perturbation (Jitter)', in *Clinical Measurements of Speech and Voice.* 2nd ed Albany, New York: Thomson Delmar Learning, pp. 190-213.

Baken, R. J. and Orlikoff, R. F. (2000c) *Clinical Measurements of Speech and Voice.* 2nd ed Albany, New York: Thomson Delmar Learning.

Baken, R. J. and Orlikoff, R. F. (2000d) 'Electroglottography', in *Clinical Measurements of Speech and Voice.* 2nd ed Albany, New York: Thomson Delmar Learning, pp. 413-427.

Behrmann, A., Agresti, C. J., Blumstein, E. and Norris, L. (1998) 'Microphone and Electroglottographic Data from Dysphonic Patients: Type 1, 2 and 3 Signals', *Journal of Voice,* 12, (2), pp. 249-260.

Bele, I. V. (2005) 'Reliability in perceptual analysis of voice quality', *Journal of Voice,* 19, (4), pp. 555-73.

Bermúdez de Alvear, R., Barón, F. and Martínez-Arquero, A. (2010) 'School Teachers' Vocal Use, Risk Factors, and Voice Disorder Prevalence: Guidelines to Detect Teachers with Current Voice Problems.', *Folia Phoniatrica et Logopaedica,* 63, (4), pp. 209-215.

Bhuta, T., Patrick, L. and Garnett, J. (2004) 'Perceptual evaluation of voice quality and its correlation with acoustic measurements.', *Journal of Voice,* 18, (3), pp. 299-304.

Bielamowicz, S., Kreimann, J., Geratt, B. R., Dauer, M. S. and Berke, B. S. (1996) 'Comparison of voice analysis systems for perturbation measurement', *Journal of Speech and Hearing Research,* 39, (1), pp. 126-134.

Bland, J. and Altman, D. (1986) 'Statistical methods for assessing agreement between two methods of clinical measurement', *Lancet,* 8, (1), pp. 307-10.

Boersma, P. (2009) 'Should jitter be measured by peak picking or by waveform matching?', *Folia Phoniatrica et Logopaedica*, 61, pp. 305-308.

Boersma, P. and Weenink, D. (2006) *PRAAT* (4.4.16, retrieved January 07, 2006, from http://www.praat.org/) Paul Boersma and David Weenink; Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016CG Amsterdam, The Netherlands.

Boersma, P. and Weenink, D. (2009) *PRAAT* (5.1.03, retrieved April 03, 2009, from http://www.praat.org/) Paul Boersma and David Weenink; Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016CG Amsterdam, The Netherlands.

Brockmann-Bauser, M. and Drinnan, M. J. (2011) 'Routine acoustic voice analysis: time to think again?', *Current Opinion in Otolaryngology, Head and Neck Surgery*, 19, (3), pp. 165-70.

Brockmann, M., Drinnan, M. J., Storck, C. and Carding, P. N. (2011) 'Reliable Jitter and Shimmer Measurements in Voice Clinics: The Relevance of Vowel, Gender, Vocal Intensity, and Fundamental Frequency Effects in a Typical Clinical Task.', *Journal of Voice*, 25, (1), pp. 44-53.

Brockmann, M., Storck, C., Carding, P. N. and Drinnan, M. J. (2008) 'Voice Loudness and Gender Effects on Jitter and Shimmer in Healthy Adults', *Journal of Speech, Language and Hearing Research*, 51, (5), pp. 1152-60.

Brown, W., Rothman, H. and Sapienza, C. (2000) 'Perceptual and acoustic study of professionally trained versus untrained voices', *Journal of Voice*, 14, (3), pp. 301-9.

Brown, W. J., Morris, R. and Murry, T. (1996) 'Comfortable effort level revisited', *Journal of Voice*, 10, (3), pp. 299-305.

Brown, W. S. and Shrivatsav, R. (2007) 'Comfortable effort level in young children's speech', *Folia Phoniatrica et Logopaedica*, 59, (5), pp. 227-233.

Carding, P. N., Steen, I. N., Webb, A., Mackenzie, K., Deary, I. J. and Wilson, J. A. (2004) 'The reliability and sensitivity to change of acoustic measures of voice quality', *Clinical Otolaryngology*, 29, pp. 538-544.

Carding, P. N., Wilson, J. A., Mackenzie, K. and Deary, I. J. (2009) 'Measuring voice outcomes: state of the science review', *The Journal of Laryngology and Otology*, 123, (8), pp. 823-9.

Casiano, R. R., Zaveri, V. and Lundy, D. S. (1992) 'Efficacy of videostroboscopy in the diagnosis of voice disorders', *Otolarygology, Head and Neck Surgery*, 107, (1), pp. 95-100.

Cheng, Y., Li, Z., Huang, J., Xue, F., Jiang, M., Wu, K. and Wang, Q. (2009) 'Combination of autologous fascia lata and fat injection into the vocal fold via the cricoid gap for unilateral vocal fold paralysis', *Archives of Otolaryngology, Head & Neck Surgery*, 135, (8), pp. 759-63.

Cho, S., Yin, C., Park, Y. and Park, Y. (2011) 'Differences in Self-Rated, Perceived, and Acoustic Voice Qualities Between High- and Low-Fatigue Groups.', *Journal of Voice,* 25, (5), pp. 544-552.

Colton, R. H., Casper, J. K. and Leonard, R. (2006) *Understanding voice problems: a physiological perspective for diagnosis and treatment* (Eds, Lappies, P., Hauber, M. and Montalbano, J.) Lippincott Williams & Wilkins, Baltimore/ Philadelphia, USA, pp. 498.

Colton, R. H. and Conture, E. G. (1990) 'Problems and pitfalls of electroglottography', *Journal of Voice,* 4, (1), pp. 10-24.

Dedivitis, R. A., Barros, A. P., Queija, D. S., Alexandre, J. C., Rezende, W. T., Corazza, V. R., Silva, V. F. and Nishimoto, I. N. (2004) 'Interobserver perceptual analysis of smokers voice', *Clinical Otolaryngology and Allied Sciences*, 29, (2), pp. 124-7.

Deem, J., Manning, W., Knack, J. and Matesich, J. (1989) 'The automatic extraction of pitch perturbation using microcomputers: some methodological considerations', *Journal of Speech and Hearing Research,* 32, pp. 689-97.

Dehgan, A., Ansari, H. and Bakhtiar, M. (2010) 'Objective voice analysis of Iranian speakers with normal voices.', *Journal of Voice,* 24, (2), pp. 161-7.

Dejonckere, P. (1998) 'Effect of louder voicing on acoustical measurements in dysphonic patients', *Logopedics, Phoniatrics, Vocology,* 23, pp. 79-84.

Dejonckere, P., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchmann, L. and Millet, B. (1996) 'Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements', *Revue de laryngologie, d'otologie et de rhinologie,* 117, (3), pp. 219-224.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchmann, L., Friedrich, G., de, P. H. v., Remacle, M. and Woisard, V. (2001) 'A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques.', *European Archives of Otorhinolarygology,* 258, pp. 77-82.

Dejonckere, P. H., Crevier-Buchmann, L., Marie, J. P., Moerman, M., Remacle, M. and Woisard, V. (2003) 'Implementation of the European Laryngological Society (ELS) basic protocol for assessing voice treatment effect.', *Revue de Laryngologie Otologie Rhinologie,* 124, (5), pp. 279-83.

Deliyski, D. and Hillman, R. (2010) 'State of the art laryngeal imaging: research and clinical implications', *Current Opinion in Otolaryngology, Head and Neck Surgery,* 18, pp. 147-152.

Deliyski, D., Shaw, H., Evans, M. and Vesselinov, R. (2006) 'Regression tree approach to studying factors influencing acoustic voice analysis', *Folia Phoniatrica et Logopaedica,* 58, pp. 274-288.

Dixon, H. (1999) 'Dysphonia and delayed food allergy: a provocation/ neutralization study with videolaryngostroboscopy', *Otology, Head and Neck Surgery,* 121, (418-429).

Dwire, A. and McCauley, R. (1995) 'Repeated measures of vocal fundamental frequency perturbation obtained using the Visi-Pitch', *Journal of Voice,* 9, (2), pp. 156-62.

Eksteen, E., Rieger, J., Nesbitt, M. and Seikaly, H. (2003) 'Comparison of voice characteristics following three different methods of treatment for laryngeal cancer', *Journal of Otolaryngology*, 32, (4), pp. 250-3.

Eskenazi, L., D.G., C. and Hicks, D. M. (1990) 'Acoustic correlates of vocal quality', *Journal of Speech and Hearing Research,* 33, (2), pp. 298-306.

Estis, J. M., Dean-Claytor, A., Moore, R. E. and Rowell, T. L. (2011) 'Pitch-Matching Accuracy in Trained Singers and Untrained Individuals: The Impact of Musical Interference and Noise' *Journal of Voice*, 25, (2), pp. 173-180.

Fant, G. (1980) 'The relations between area functions and the acoustic signal', *Phonetica,* 37, (1-2), pp. 55-86.

Fitch, J. L. (1990) 'Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech.', *Journal of Speech & Hearing Disorders,* 55, (2), pp. 360-3.

Fourcin. *Fourcin Laryngograph Processor,* Laryngograph Ltd., Wallington GB.

Fukazawa, T., Blaugrund, S., El-Assuooty, A. and Gould, W. (1988) 'Acoustic analysis of hoarse voice : a preliminary report', *Journal of Voice*, 2, pp. 127-32.

Gelfer, M. P. (1995) 'Fundamental Frequency, Intensity, and Vowel Selection: Effects on Measures of Phonatory Stability', *Journal of Speech and Hearing Research,* 38, pp. 1189-1198.

Gelfer, M. P. and Fendel, D. (1995) 'Comparisons for Jitter, Shimmer, Signal-to Noise Ratio from Directly Digitized Versus Taped Voice Samples', *Journal of Voice,* 9, (4), pp. 378-382.

Gould, J., Waugh, J., Carding, P. N. and Drinnan, M. J. (2011) 'A new voice rating tool for clinical practice', *Journal of Voice,* Dec 29 epub ahead of print.

Gramming, P. (1988) 'Relationship between changes in voice pitch and loudness', *Journal of Voice,* 2, pp. 118-126.

Haji, T., Horiguchi, S., Baer, T. and Gould, W. (1986) 'Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation.', *Journal of the Acoustical Society of America,* 80, (1), pp. 58-62.

Hanson, D. G., Jiang, J. J., Chen, J. and Pauloski, B. R. (1997) 'Acoustic measurement of change in voice quality with treatment for chronic posterior laryngitis', *Annals of Otology, Rhinology and Laryngology,* 106, (4), pp. 279-285.

Hartl, D., Hans, S., Crevier-Buchmann, L., Vaissière, J. and Brasnu, D. (2009) 'Long-term acoustic comparison of thyroplasty versus autologous fat injection.', *Annals of Otology, Rhinology and Laryngology,* 118, (12), pp. 827-32.

Heilberger, V. and Horii, Y. (1982) 'Jitter and Shimmer in sustained phonation', in Lass, N.(ed), *Speech and language: advances in basic research and practice.* Vol. 7 New York: Academic Press, pp. 299-332.

Heylen, L., Wuyts, F. L., Mertens, P., De Bodt, M. S. and Van de Heyning, P. H. (2002) 'Normative voice range profiles of male and female professional voice users', *Journal of Voice*, 16, (1), pp. 1-7.

Hirano, M. (1974) 'Morphological structure of the vocal cord as a vibrator and its variations', *Folia Phoniatrica*, 26, (2), pp. 89-94.

Hirano, M. (1981) *Clinical Examination of Voice.* Wien /New York: Springer-Verlag.

Hirose, H. (1999) 'Investigating the Physiology of Laryngeal Structures', in Hardcastle, W. and Laver, J.(eds) *The Handbook of Phonetic Sciences.* 2 ed Oxford UK: Blackwell Publishers Ltd., pp. Page 126, Figure 4.6.

Hodge, F., Colton, R. and Kelley, R. (2001) 'Vocal intensity characteristics in normal and elderly speakers', *Journal of Voice*, 15, (4), pp. 503-511.

Honda, K. (1983) 'Relationship between pitch control and vowel articulation', in Bless, D. M. and Abbs, J. H.(eds) *Vocal Fold Physiology: Contemporary Research and Clinical Issues.* San Diego: College-Hill Press, pp. 286-297.

Horii, Y. (1980) 'Vocal Shimmer in sustained phonation', *Journal of Speech and Hearing Research*, 23, pp. 202-09.

Jacobson, B., Johnson, A. and Grywalsky, C. (1996) 'The Voice Handicap Index (VHI): development and validation', *American Journal of Speech-Language Pathology,* 6, pp. 66-70.

Jafari, M., Till, J., Truesdell, L. and Law-Till, C. (1993) 'Time-shift, trial and gender effects on vocal perturbation measurements', *Journal of Voice,* 7, pp. 326-36.

Jiang, J., Zhang, Y., MacCallum, J., Sprecher, A. and Zhou, L. (2009) 'Objective acoustic analysis of pathological voices from patients with vocal nodules and polyps.', *Folia Phoniatrica et Logopaedica,* 61, (6), pp. 342-9.

Jones, S., Carding, P. and Drinnan, M. (2006) 'Exploring the relationship between severity of dysphonia and voice-related quality of life', *Clinical Otolaryngology,* 31, (5), pp. 411-417.

Jonsdottir, V., Laukkanen, A. and Siikki, I. (2003) 'Changes in teacher`s voice quality during a working day with and without electric sound amplification', *Folia Phoniatrica et Logopaedica,* 55, (5), pp. 267-80.

Karnell, M. (1991) 'Laryngeal perturbation analysis: minimum length of analysis window', *Journal of Speech and Hearing Research,* 34, pp. 544-48.

Karnell, M., Hall, K. D. and Landahl, K. L. (1995) 'Comparison of Fundamental Frequency and Perturbation Measurements Among Three Analysis Systems', *Journal of Voice,* 9, (4), pp. 383-393.

Kay Elemetrics Multi-Dimensional Voice Program (1993), KayPentax, Lincoln Park NJ, USA.

Kent, R. D. (1993) 'Vocal tract acoustics', *Journal of Voice*, 7, (2), pp. 97-117.

Kent, R. D., Kent, J. F. and Rosenbek, J. C. (1987) 'Maximum performance tests of speech production', *The Journal of Speech and Hearing Disorders*, 52, (4), pp. 367-87.

Kiliç, M. A., Ögüt, F., Dursun, G., Okur, E., Yildirim, I. and Midilli, R. (2004) 'Effects of Vowels on Voice Perturbation Measures', *Journal of Voice*, 18, (3), pp. 318-324.

Kleemola, L., Helminen, M., Rorarius, E., Isotalo, E. and Sihvo, M. (2011) 'Voice activity and participation profile in assessing the effects of voice disorders on quality of life: estimation of the validity, reliability and responsiveness of the finnish version.', *Folia Phoniatrica et Logopaedica*, 63, (3), pp. 113-121.

Kreiman, J. and Geratt, B. R. (2000) 'Sources of listener disagreement in voice quality assessment', *Journal of the Acoustical Society of America*, 108, (4), pp. 1867-76.

Kreiman, J. and Geratt, B. R. (2007) 'When and why listeners disagree in voice quality assessment tasks.', *Journal of the Acoustical Society of America*, 122, (4), pp. 2354-64.

Kreiman, J., Geratt, B. R., Kempster, G. B., Erman, A. and Berke, B. S. (1993) 'Perceptual evaluation of voice quality: review, tutorial, and a framework for future research', *Journal of Speech and Hearing Research,* 36, (1), pp. 21-40.

Ladefoged, P. and Maddieson, I. (1996) 'Places of Articulation', in *The sounds of the world's languages. Phonological Theory.* 1st ed Oxford: Wiley-Blackwell, pp. 9-15.

Landis, J. and Coch, G. (1977) 'The measurement of observer agreement for categorical data', *Biometrics,* 33, (1), pp. 159-174.

Laukkanen, A., Ilomäki, I., Leppänen, K. and Vilkman, E. (2008) 'Acoustic measures and self-reports of vocal fatigue by female teachers.', *Journal of Voice*, 22, (3), pp. 283-9.

Laukkanen, A. and Kankare, E. (2006) 'Vocal loading-related changes in male teachers' voices investigated before and after a working day.', *Folia Phoniatrica et Logopaedica,* 58, (4), pp. 229-39.

Lee, S., Kim, J., Koh, Y., Shim, S. and Son, Y. (2010) 'Comparative Analysis of Efficiency of Injection Laryngoplasty Technique for with or without Neck Treatment Patients: A Transcartilaginous Approach Versus the Cricothyroid Approach.', *Clinical & Experimental Otolaryngology,* 3, (1), pp. 37-41.

Licitra, L., Bernier, J., Grandi, C., Locati, L., Merlano, M., Gatta, G. and Lefebvre, J. (2003) 'Cancer of the larynx', *Oncology Hematology,* 47, pp. 65-80.

Liebermann, P. (1963) 'Some acoustic measures of the fundamental periodicity of normal and pathologic larynxes', *Journal of the Acoustical Society of America,* 35, pp. 344-53.

Lin, E., Jiang, J., Noon, S. and Hanson, D. G. (2000) 'Effects of head extension and tongue protrusion on voice perturbation measures', *Journal of Voice,* 14, (1), pp. 8-16.

Lindstrom, F., Ohlsson, A., Sjöholm, J. and Waye, K. (2010) 'Mean F0 values obtained through standard phrase pronunciation compared with values obtained from the normal work environment: a study on teacher and child voices performed in a preschool environment.', *Journal of Voice,* 24, (3), pp. 319-23.

LingWAVES Vospector (2006) LingCom/ Wevosys, Forchheim, Germany.

Linville, S. E. and Korabic, E. W. (1987) 'Fundamental frequency stability characteristics of elderly women's voices', *Journal of the Acoustical Society of America,* 81, (4), pp. 1196-9.

Ludlow, C., Bassich, C., Connor, N., Connor, N., Coulter, D. and Lee, Y. (1987) 'The validity of using jitter and shimmer to detect laryngeal pathology', in Baer, T., Sasaki, C. and Harris, K.(eds) *Laryngeal function in phonation and respiration: vocal fold physiology series.* Boston: College Hill, pp. 492-508.

Ma, E. and Yiu, E. (2005) 'Suitability of acoustic perturbation measures in analysing periodic and nearly periodic voice signals', *Folia Phoniatrica et Logopaedica,* 57, (1), pp. 38-47.

Ma, E. and Yiu, E. (2006) 'Multiparametric evaluation of dysphonic severity', *Journal of Voice*, 20, (3), pp. 380-90.

Maccallum, J., Zhang, Y. and Jiang, J. (2010) 'Vowel Selection and Its Effects on Perturbation and Nonlinear Dynamic Measures', *Folia Phoniatrica et Logopaedica*, 63, (2), pp. 88-97.

Martens, J., Versnel, H. and Dejonckere, P. (2007) 'The effect of visible speech in the perceptual rating of pathological voices.', *Archives of Otolaryngology, Head & Neck Surgery*, 133, (2), pp. 178-85.

Maryn, Y., Corthals, P., Bodt De, M., Cauwenberge van, P. and Delyiski, D. (2009) 'Perturbation measures of voice: a comparative study between Multi-Dimensional Voice Program and Praat', *Folia Phoniatrica et Logopaedica*, 61, (4), pp. 217-226.

Mazzoni, D., Brubeck, M., Crook, J., Johnson, V., Meyer, M., Habermann, J., Montgomery, M. and Mueller, S. (2005) *Audacity* (1.2.4b)

Mazzoni, D., Brubeck, M., Crook, J., Johnson, V., Meyer, M., Habermann, J., Montgomery, M. and Mueller, S. (2008) *Audacity* (1.2.6)

Mehta, D., Deliyski, D., Zeitels, S., Quatieri, T. and Hillman, R. (2010) 'Voice Production Mechanisms Following Phonosurgical Treatment of Early Glottic Cancer', *Annals of Otology, Rhinology and Laryngology*, 119, pp. 1-9.

Mehta, D. and Hillman, R. (2008) 'Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods.', *Current Opinion in Otolaryngology, Head and Neck Surgery,* 16, (3), pp. 211-5.

Milenkovic, P. (1987) 'Least mean square measures of voice perturbation', *Journal of Speech and Hearing Research,* 30, pp. 529-38.

Milenkovic, P. and Read, C. (1992) *CSpeech*, University of Wisconsin-Madison, USA.

Murry, T. and Doherty, E. T. (1980) 'Selected Acoustic Characteristics of Pathologic and Normal Speakers', *Journal of Speech and Hearing Research,* 23, pp. 361-369.

Nawka, T., Wiesmann, U. and Gonnermann, U. (2003) 'Validierung des Voice Handicap Index in der deutschen Fassung (German: Validation of the german version of the Voice Handicap Index)', *HNO,* 51, pp. 921-929.

Niebudek-Bogusz, E., Kotylo, P., Politanski, P. and Sliwinska-Kowalska, M. (2008) 'Acoustic analysis with vocal loading test in occupational voice disorders: outcomes before and after voice therapy.', *International Journal of Occupational Medicine & Environmental Health,* 21, (4), pp. 301-8.

Niebudek-Bogusz, E., P., K. and Sliwinska-Kowalska, M. (2007) 'Evaluation of voice acoustic parameters related to the vocal-loading test in professionally active teachers with dysphonia.', *International Journal of Occupational Medicine & Environmental Health*, 20, (1), pp. 25-30.

Niebudek-Bogusz, E., Woznicka, E., Zamyslowska-Szmytke, E. and Sliwinska-Kowalska, M. (2010) 'Correlation between acoustic parameters and Voice Handicap Index in dysphonic teachers.', *Folia Phoniatrica et Logopaedica*, 62, (1-2), pp. 55-60.

Nittrouer, S., McGowan, R. S., Milenkovic, P. H. and Beehler, D. (1990) 'Acoustic Measurements of Men's and Women's Voices. A Study of Context Effects and Covariation', *Journal of Speech and Hearing Research*, 33, pp. 761-775.

Orlikoff, R. (1995) 'Vocal Stability and Vocal Tract Configuration: An Acoustic Electroglottographic Investigation', *Journal of Voice*, 9, (2), pp. 173-181.

Orlikoff, R. F. and Baken, R. (1990) 'Consideration of the relationship between the Fundamental frequency of phonation and vocal Jitter', *Folia Phoniatrica*, 42, pp. 31-40.

Orlikoff, R. F. and Baken, R. J. (1989) 'The effect of the heartbeat on vocal fundamental frequency perturbation.', *Journal of Speech and Hearing Research*, 32, (3), pp. 576-82.

Orlikoff, R. F. and Kahane, J. C. (1991) 'Influence of mean Sound Pressure Level on Jitter and Shimmer measures', *Journal of Voice*, 5, (2), pp. 113-119.

Pabon, J. (1991) 'Objective acoustic voice-quality parameters in the computer Phonetogram', *Journal of Voice*, 5, (3), pp. 203-216.

Pabon, J. P. and Plomp, R. (1988) 'Automatic phonetogram recording supplemented with acoustical voice-quality parameters', *Journal of Speech and Hearing Research*, 31, (4), pp. 710-22.

Perry, C., Ingrisano, D.-S., Palmer, M. and McDonald, E. (2000) 'Effects of Environmental Noise on Computer-derived Voice Estimates from Female Speakers', *Journal of Voice,* 14, (2), pp. 146-153.

Pfüetzenreiter, E. J., Dedivitis, R., Queija, D., Bohn, N. and Barros, A. (2010) 'The relationship between the glottic configuration after frontolateral laryngectomy and the acoustic voice analysis.', *Journal of Voice,* 24, (4), pp. 499-502.

Porter, M. T. (1995) *Trust in numbers: the pursuit of objectivity.* Princeton, New Jersey: Princeton University Press.

Pribuisiene, R., Uloza, V., Kupcinskas, L. and Jonaitis, L. (2006) 'Perceptual and acoustic characteristics of voice changes in reflux laryngitis patients', *Journal of Voice,* 20, (1), pp. 128-36.

Rabinov, C. R., Kreimann, J., Geratt, B. R. and Bielamowicz, S. (1995) 'Comparing Reliability of Perceptual Ratings of Roughness and Acoustic Measures of Jitter', *Journal of Speech and Hearing Research,* 38, pp. 26-32.

Ramig, L.O., Ringel, R.L. (1983) 'Effects of physiological aging on selected acoustic characteristics of voice', *Journal of Speech and Hearing Research,* 26, pp. 22-30.

Rantala, L., Paavola, L., Korkko, P. and Vilkman, E. (1998) 'Working-day effects on the spectral characteristics of teaching voice', *Folia Phoniatrica et Logopaedica,* 50, pp. 205-211.

Rantala, L. and Vilkman, E. (1999) 'Relationship between subjective voice complaints and acoustic parameters in female teachers' voices.', *Journal of Voice,* 13, (4), pp. 484-495.

Rantala, L., Vilkman, E. and Bloigu, R. (2002) 'Voice changes during work: subjective complaints and objective measurements for female primary and secondary schoolteachers.', *Journal of Voice,* 16, (3), pp. 344-55.

Rosen, C. A., Lombard, L. E. and Murry, T. (2000) 'Acoustic, aerodynamic, and videostroboscopic features of bilateral vocal fold lesions.', *Annals of Otology, Rhinology and Laryngology,* 109, (9), pp. 823-8.

Rosen, C. A. and Simpson, C. B. (2008) *Operative Techniques in Laryngology*. Berlin, Heidelberg: Springer-Verlag, Germany.

Rovirosa, A., Martinez-Celdran, E., Ortega, A., Ascaso, C., Abellana, R., Velasco, M., Bonet, M., Herrera, C., Casas, F., Francisco, R., Arenas, M., Hernandez, V., Sanchez-Reyes, A., Leon, C., Traserra, J. and Biete, A. (2000) 'Acoustic analysis after radiotherapy in T1 vocal cord carcinoma: a new approach to the analysis of voice quality', *International Journal of Radiation Oncology, Biology, Physics,* 47, (1), pp. 73-9.

Roy, N., Merrill, R., Thibeault, S., Parsa, R., Gray, S. and Smith, E. (2004) 'Prevalence of voice disorders in teachers and the general population.', *Journal of Speech, Language and Hearing Research,* 47, (2), pp. 281-93.

Roy, N., Weinrich, B., Gray, S., Tanner, K., Toledo, S., Dove, H., Corbin-Lewis, K. and Stemple, J. (2002) 'Voice amplification versus vocal hygiene instruction for teachers with voice disorders: a treatment outcomes study', *Journal of Speech Language & Hearing Research,* 45, (4), pp. 625-38.

Ruas, A., Rolla, V., Araújo-Melo de, M., Moreira, J. and Valete-Rosalino, C. (2010) 'Vocal quality of patients treated for laryngeal tuberculosis, before and after speech therapy.', *The Journal of Laryngology and Otology,* 124, (11), pp. 1153-7.

Sama, A., Carding, P. N., Price, S., Kelly, P. and Wilson, J. A. (2001) 'The Clinical Features of Functional Dysphonia', *Laryngoscope,* 111, (3), pp. 458-463.

Sanuki, T., Yumoto, E., Minoda, R. and Kodama, N. (2010) 'Effects of type II thyroplasty on adductor spasmodic dysphonia.', *Otolarygology, Head and Neck Surgery,* 142, (4), pp. 540-6.

Scherer, R. C. (1995) 'Required number of tokens to determine representative voice perturbation values', *Journal of Speech and Hearing Research,* 38, pp. 1260-1269.

Schindler, A., Mozzanica, F., Vedroy, M., Maruzzi, P. and Ottaviani, F. (2009) 'Correlation between the Voice Handicap Index and voice measurements in four groups of patients with dysphonia', *Otolarygology, Head and Neck Surgery,* 141, pp. 762-769.

Schmidt, C. P., Andrews, M. L. and McCutcheon, W. (1998) 'An acoustical and perceptual analysis of the vocal behaviour of classroom teachers', *Journal of Voice,* 12, (4), pp. 434-43.

Schneider, B., Wendler, J. and W., S. (2002) 'The relevance of stroboscopy in functional dysphonias', *Folia Phoniatrica,* 54, pp. 44-54.

Schwartz, S. R., Cohen, S.M., Dailey, S.H., Rosenfeld, R.M., Deutsch, E.S., Boyd Gillespie, M., Granierei, E., Hapner, E.R., Kimball, E., Krouse, H.J., McMurray, J.S., Medina, S., O´Brien, K., Ouellette, D.R., Messinger-Rapport, B.J., Stachler, R.J., Strode, S., Thompson, D.M., Stemple, J.C., Willging, P.J., Cowley, T., McCoy, S., Bernad, P.G. and Patel, M.M. (2009) 'Efficacy of videostroboscopy in the diagnosis of voice disorders', *Otolarygology, Head and Neck Surgery,* 141, (3), pp. 1-31.

Schoentgen, J. (1982) 'Qualitative evaluation of the discriminating performance of acoustic features in detecting laryngeal pathology', *Journal of Speech and Hearing Research,* 13, pp. 489-502.

Shah, R., Engel, S. and Choi, S. (2008) 'Relationship between voice quality and vocal nodule size', *Otolarygology, Head and Neck Surgery,* 139, (5), pp. 723-6.

Shao, J., MacCallum, J., Zhang, Y., Sprecher, A. and Jiang, J. (2010) 'Acoustic analysis of the tremulous voice: assessing the utility of the correlation dimension and perturbation parameters.', *Journal of Communication Disorders,* 43, (1), pp. 35-44.

Sjögren, E. V., Rossum van, M. A., Langeveld, T. P., Voerman, M. S., Kamp van de, V. A., Friebel, M. O., Wolterbeek, R. and Baatenburg de Jong, R. J. (2008) 'Voice outcome in T1a midcord glottic carcinoma: laser surgery vs radiotherapy.', *Archives of Otolaryngology, Head & Neck Surgery,* 134, (9), pp. 965-72.

Sliwinska-Kowalska, M., Niebudek-Bogusz, E., Fiszer, M., Los-Spychalska, T., Kotylo, P., Sznurowska-Przygocka, B. and Modrzewska, M. (2006) 'The prevalence and risk factors for occupational voice disorders in teachers.', *Folia Phoniatrica*, 58, (2), pp. 85-101.

Sobotta. (2006) 'Hals (Neck)', in Putz, R. and Pabst, R.(eds) *Atlas der Anantomie des Menschen (Atlas of human anatomy).* Vol. 1 Munich, Jena: Urban&Fischer, pp. 123-131.

Sorensen, D. and Horii, Y. (1983) 'Frequency and amplitude perturbation in the voices of female speakers', *Journal of Communication Disorders,* 16, pp. 57-61.

Stathopoulos, E., Huber, J. and Sussmann, J. (2011) 'Changes in Acoustic Characteristics of the Voice across the Life-span: Measures from 4-93 Year Olds.', *Journal of Speech, Language and Hearing Research,* 54, (4), pp. 1011-1021.

Stemple, J. C., Glaze, L. E. and Gerdeman, B. K. (2000) *Clinical voice pathology: theory and management.* 3 ed San Diego (CA): Delmar. Singular Publishing Group.

Stojadinovic, A., Shaha, A. R., Orlikoff, R. F., Nissan, A., Kornak, M. F., Singh, B., Boyle, J. O., Shah, J. P., Brennan, M. F. and Kraus, D. H. (2002) 'Prospective functional voice assessment in patients undergoing thyroid surgery', *Annals of surgery*, 236, (6), pp. 554-60.

Sulter, A. M., H.K.Schutte and Miller, D. G. (1996) 'Standardized laryngeal videostroboscopic rating: differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age', *Journal of Voice,* 10, (2), pp. 175-189.

Sussmann, J. E. and Sapienza, C. (1994) 'Articulatory, Developmental, and Gender Effects on Measures of Fundamental Frequency and Jitter', *Journal of Voice,* 8, (2), pp. 145-156.

Titze, I. R. (1988) 'The physics of small-amplitude oscillation of the vocal folds', *Journal of the Acoustical Society of America,* 83, (4), pp. 1536-1552.

Titze, I. R. (1991) 'A model for neurologic sources of aperiodicity in vocal fold vibration', *Journal of Speech and Hearing Research,* 34, pp. 460-72.

Titze, I. R. (1995) *Workshop on Acoustic Voice Analysis: Summary Statement,* Iowa City: National Center for Voice and Speech.

Titze, I. R., Horii, Y. and Scherer, R. C. (1987) 'Some technical considerations in voice perturbation measurements', *Journal of Speech & Hearing Research,* 30, (2), pp. 252-60.

Titze, I. R. and Liang, H. (1993) 'Comparison of F0 Extraction Methods for High Precision Voice Perturbation Measurements', *Journal of Speech and Hearing Research*, 36, pp. 1120-1133.

Van Houtte, E., Claeys, S., Wuyts, F. and Van Lierde, K. (2011) 'The Impact of Voice Disorders Among Teachers: Vocal Complaints, Treatment-Seeking Behavior, Knowledge of Vocal Care, and Voice-Related Absenteeism.', *Journal of Voice,* 25, (5), pp. 570-5.

Van Houtte, E., Van Lierde, K., D`Haeseleer, E. and Claeys, S. (2010) 'The prevalence of laryngeal pathology in a treatment-seeking population with dysphonia.', *Laryngoscope,* 120, (2), pp. 306-12.

Van Lierde, K. M., De Bodt, M. S., Dhaeseleer, E., Wuyts, F. and Claeys, S. (2010) 'The treatment of muscle tension dysphonia: a comparison of two treatment techniques by means of an objective multiparameter approach.', *Journal of Voice,* 24, (3), pp. 294-301.

Vashani, K., Murugesh, M., Hattiangadi, G., Gore, G., Keer, V., Ramesh, V., Sandur, V. and SJ, B. (2010) 'Effectiveness of voice therapy in reflux-related voice disorders', *Diseases of the Esophagus,* 23, (1), pp. 27-32.

Vieira, M. N., McInnes, F. R. and Jack, M. A. (2002) 'On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures.', *Journal of the Acoustical Society of America,* 111, (2), pp. 1045-55.

Wagener, F., Kühnel, V. and Kollmeier, B. (1999) 'Entwicklung und Evaluation eines Satztestes in deutscher Sprache I: Design des Oldenburger Satztestes (German) "Development and evaluation of a sentence based hearing test in german language I: Design of the Oldenburger Satztest."', *Zeitschrift für Audiologie,* 38, (1), pp. 4-15.

Wang, J. and Jo, C. (2007) *Engineering in Medicine and Biology Society, 2007 (EMBS 2007); 29th Annual International Conference of the IEEE.* Lyon, 22.-26.08.2007.The Printing House Inc.

Watts, C., Moore, R. and McCaghren, K. (2005) 'The Relationship Between Vocal Pitch-Matching Skills and Pitch Discrimination Skills in Untrained Accurate and Inaccurate Singers' *Journal of Voice,* **19,** 534-543.

Wilcox, K. and Horii, Y. (1980) 'Age and changes in vocal jitter', *Journal of Gerontology,* 35, pp. 194-8.

Winholtz, W. and Titze, I. (1997a) 'Conversion of a head-mounted microphone signal into calibrated SPL units', *Journal of Voice,* 11, (4), pp. 417-421.

Winholtz, W. and Titze, I. R. (1997b) 'Miniature Head-Mounted Microphone for Voice Perturbation Analysis', *Journal of Speech, Language and Hearing Research,* 40, (894-899).

Wolfe, V. and Martin, D. (1997) 'Acoustic correlates of dysphonia: type and severity.', *Journal of Communication Disorders,* 30, (5), pp. 403-415.

Wuyts, F. L., Bodt De, M., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., Lierde Van, K. and Raes, J. (2000) 'The Dysphonia Severity Index. An Objective Measure of Vocal Quality Based on a Multiparameter Approach', *Journal of Speech, Language and Hearing Research,* 43, pp. 796-809.

Yelken, K., Gultekin, E., Guven, M., Eyibilen, A. and Aladag, I. (2010) 'Impairment of voice quality in paradoxical vocal fold motion dysfunction.', *Journal of Voice,* 24, (6), pp. 724-7.

Zagólski, O. (2009) 'Electroglottography in Elderly Patients with Vocal-Fold Palsy', *Journal of Voice,* 23, (5), pp. 567-571.

Zhang, Y., Sprecher, A., Wei, C. and Jiang, J. (2010) 'Implantation of gelatin sponge combined with injection of autologous fat for sulcus vocalis.', *Otolarygology, Head and Neck Surgery,* 143, (2), pp. 198-203.

Zyski, B., Bull, G., McDonald, W. and Johns, M. (1984) 'Perturbation analysis of normal and pathological larynges', *Folia Phoniatrica,* 36, pp. 190-8.

Appendices

Appendix A: Publishing details and statement of contribution

Chapter 2

Brockmann-Bauser, M. and Drinnan, M. J. (2011) 'Routine acoustic voice analysis: time to think again?', *Current Opinion in Otolaryngology, Head and Neck Surgery,* 19, (3), pp. 165-70.

Statement of contribution: M. Brockmann-Bauser developed, wrote and revised the article text. M. Drinnan provided the data and statistical analysis for Figure 1 and Figure 2. Also he revised the text in terms of structure, scientific content and language.

Chapter 4

Brockmann, M., Storck, C., Carding, P. N. and Drinnan, M. J. (2008) 'Voice Loudness and Gender Effects on Jitter and Shimmer in Healthy Adults', *Journal of Speech, Language and Hearing Research,* 51, (5), pp. 1152-60.

Statement of contribution: M. Brockmann designed the study together with M. Drinnan, P. Carding and C. Storck. M. Brockmann collected the data, prepared the voice recordings and conducted the acoustic analysis including quality control. Further, she wrote and revised the article text. C. Storck assisted the data collection, conducted perceptual assessments and gave regular feedback on the article text. P. Carding revised the article text in terms of structure, scientific content and language. M. Drinnan advised on measurement technique, performed the statistical analysis and revised the text in terms of structure, scientific content and language.

Chapter 5

Brockmann, M., Drinnan, M. J., Storck, C. and Carding, P. N. (2011) 'Reliable Jitter and Shimmer Measurements in Voice Clinics: The Relevance of Vowel, Gender, Vocal Intensity, and Fundamental Frequency Effects in a Typical Clinical Task.', *Journal of Voice*, 25, (1), pp. 44-53.

Statement of contribution: M. Brockmann designed the study together with M. Drinnan, P. Carding and C. Storck. M. Brockmann collected the data, prepared the voice recordings and conducted the acoustic analysis including quality control. Further, she wrote and revised the article text. C. Storck assisted the data collection, conducted perceptual assessments and gave regular feedback on the text. M. Drinnan programmed the PRAAT script for data analysis, performed the inferential statistical analysis, and revised the text in terms of structure, scientific content and language. P. Carding revised the text in terms of structure, scientific content and language.

Chapter 6

Brockmann, M., Drinnan, M. J. and Carding, P. N. (2009) 'Comparison of voice intensity effects on electroglottographic versus acoustic jitter and shimmer.', *Journal of the Acoustical Society of America*, 125, (4), p. 2532 (Abstract).

Statement of contribution: M. Brockmann designed the study together with M. Drinnan and P. Carding. M. Brockmann collected the data, prepared the acoustic and electroglottographic recordings and conducted the instrumental analysis including quality control. Further, she wrote and revised the abstract text. M. Drinnan advised on the measurement technique, programmed the PRAAT script for data analysis and revised the statistical analysis. Both M. Drinnan and P. Carding revised the abstract in terms of structure, scientific content and language.

Appendix B: Paper “Voice loudness and Gender Effects on Jitter and Shimmer in Healthy Adults”

Paper summarising the preceding MSc study from M. Brockmann “Voice Loudness and gender influence on Jitter and Shimmer in healthy adults” (submitted to Newcastle University August 2006).

Voice Loudness and Gender Effects on Jitter and Shimmer in Healthy Adults

Meike Brockmann
Claudio Storck
University Hospital Zurich, Switzerland

Paul N. Carding
Michael J. Drinnan
Freeman Hospital, Newcastle upon Tyne, Great Britain

Purpose: The aim of this study was to investigate voice loudness and gender effects on jitter and shimmer in healthy young adults because previous descriptions have been inconsistent.
Method: Fifty-seven healthy adults (28 women, 29 men) aged 20–40 years were included in this cross-sectional single-cohort study. Three phonations of /a/ at soft, medium, and loud individual loudness were recorded and analyzed using PRAAT software (P. Boersma & D. Weeninkk, 2006). Voice loudness and gender effects on measured sound pressure level, fundamental frequency, jitter, and shimmer were assessed through the use of descriptive and inferential (analysis of variance) statistics.
Results: Jitter and shimmer significantly increased with decreasing voice loudness, especially in phonations below 75 dB and 80 dB. In soft and medium phonation, men were generally louder and showed significantly less shimmer. However, men had higher jitter measures when phonating softly. Gender differences in jitter and shimmer at medium loudness may be mainly linked to different habitual voice loudness levels.
Conclusion: This pragmatic study shows significant voice loudness and gender effects on perturbation. In clinical assessment, requesting phonations above 80 dB at comparable loudness between genders would enhance measurement reliability. However, voice loudness and gender effects in other age groups, in disordered voices, or when a minimal loudness is requested should be further investigated.

KEY WORDS: voice loudness, gender, jitter, shimmer, acoustic assessment

In voice clinics and research, voice perturbation analysis is considered an easily applicable, noninvasive, and inexpensive measurement of vocal output that may complement other laryngeal diagnostic methods (such as videolaryngostroboscopy). Voice perturbation parameters, such as jitter and shimmer, are routinely measured in acoustic voice assessment as recommended by European and American research associations (Dejonckere et al., 2001; Titze, 1995). These measures indirectly assess laryngeal function by quantifying acoustic correlates of irregular vocal fold vibration. *Fundamental frequency* (F_0) describes human voice pitch and indicates the number of vocal fold vibratory cycles per second (Hz). *Sound pressure level* (SPL) measures voice intensity in decibels (dB). *Jitter* measures F_0 perturbation, and *shimmer* measures SPL perturbation, caused by vibratory variations from one vocal fold cycle to the next.

Voice perturbation analysis has been widely used in diagnostics, voice treatment documentation, and characterization of acoustic voice properties of specific groups such as elderly people (Hodge, Colton, & Kelley, 2001). Jitter and shimmer also appear to relate to perceived voice "roughness" and "hoarseness" (Dejonckere et al., 1996). These parameters have also been shown to be greater in disordered voices and to discriminate between healthy and pathologic voices in some voice disorder types (Askenfelt & Hammarberg, 1986; Schoentgen, 1982). Acoustic analysis has been widely used to document voice changes following voice therapy (Roy et al., 2002) and medical, irradiation, or surgical treatment (Hanson, Jiang, Chen, & Pauloski, 1997; Rovirosa et al., 2000). Perturbation analysis has also been used to compare benefits of different intervention types (Eksteen, Rieger, Nesbitt, & Seikaly, 2003).

Limits in Application

Despite its ubiquitous use, acoustic signal perturbation measurements appear to have significant limitations. For example, Zyski, Bull, McDonald, and Johns (1984) described limited sensitivity and specificity when investigating and comparing four different jitter and shimmer measurement types. Their study showed that between 21% and 77% of mean values in voice-disordered adults were within the normal voice range. Carding et al. (2004) also reported poor to moderate retest reliability and poor sensitivity to change. Despite the use of the same jitter (jitter %) and shimmer (shimmer dB) types, reported normative values vary considerably between studies, thus hindering comparisons (see Table1). Potentially substantial SPL and gender influences have been previously identified but characterized incompletely or inconsistently.

The Influence of Voice Loudness and Gender

Voice loudness. Orlikoff and Kahane (1991) describe an inverse linear relationship of jitter and shimmer to SPL. Their data were obtained from 10 healthy men phonating at three different prescribed intensity-level ranges (60–68 dB, 70–78 dB, and 80–88 dB). Dejonckere (1998) reported a significant perturbation reduction in louder voices when comparing "comfortable loudness and pitch" phonation with "louder" phonation in 55 dysphonic adults with functional dysphonia or superficial vocal fold pathology. Similarly, other authors (Pabon, 1991; Pabon & Plomp, 1988) have reported higher jitter and shimmer values at low frequency and SPLs. Despite the consistent description of higher voice perturbation in lower SPL, the present characterization remains incomplete because (a) jitter and shimmer have been averaged for large or undocumented SPL ranges, (b) acoustic parameters and statistical tests have not been described, and/or (c) groups have been small and highly inconsistent.

Table 1. Reported mean shimmer (dB) and jitter(%) values in gender.

Authors	Participants	Shimmer (dB)	Jitter (%)
(Wilkox & Horii, 1980)	20 men	0.47	None
(Horii, 1980)	31 men	0.47	0.61
		SD = 0.34	*SD* = 0.20
(Horii, 1982)	20 men	0.13	0.66
		SD = 0.23	*SD* = 0.18
(Sorensen & Horii, 1983)	20 women	0.33	0.71
		SD = 0.22	*SD* = 0.22
(Jafari et al., 1993)	5 men	0.16–0.22	0.20–0.22
	5 women	0.12–0.14	0.24–0.34
(Ferrand, 1995)	30 women	0.19–0.21	0.446–0.674

Gender. Evidence regarding gender effects in jitter and shimmer values is also inconclusive. In previous studies, women have displayed minimally less shimmer and more jitter (Deem, Manning, Knack, & Matesich, 1989; Sorensen & Horii, 1983) but also smaller absolute jitter values than men (Jafari, Till, Truesdell, & Law-Till, 1993; Ludlow et al., 1987). In contrast, similar jitter and shimmer were found in women and men at "comfortable pitch," whereas jitter was more influenced by changing F_0 in men (Orlikoff & Baken, 1990). However, these studies are limited because of (a) small numbers of participants, (b) phonation at prescribed F_0 levels, and/or (c) lack of reporting of the produced F_0 and SPL levels.

Clinical relevance. In acoustic voice assessment, patients are commonly requested to phonate "/a/ at comfortable loudness and pitch," and usually SPL is not controlled for (Dejonckere et al., 2001; Titze, 1995). Even though men have been shown to phonate louder than women in this situation and patients tend to vary voice intensity between sessions (Brown, Morris, & Murry, 1996), acoustic analysis programs do not provide SPL-corrected normative thresholds (Boersma & Weenink, 2006; Kay Elemetrics Corp., 1999). If the effects of SPL and gender are significant, failure to consider them in voice assessments results in a false picture of the acoustic and, hence, vibratory properties of the patient's vocal folds. This is especially problematic when acoustic analysis is used to measure intervention success. Furthermore, this might delay the patient's access to appropriate further diagnostics and treatment and, in the worst case, may even support false diagnoses. The assessment of gender and SPL effects requires measurements of phonations at a wide range of loudness levels in a sufficiently large and consistent group and use of adequate recording methods (Titze, 1995).

Study Aims

The aim of this study was to assess voice loudness and gender effects on jitter and shimmer in vowel phonation in vocally healthy young female and male adults.

Method

Participants

Seventy volunteers (35 women, 35 men) between the ages of 20 and 40 years were recruited from the

University Hospital Zurich for participation in the experiment. All participants gave written consent.

Exclusion criteria. Participants were excluded from the study if they presented with a hoarse voice as perceived by the examiner on the day of recording. They were also excluded if they (a) reported recent voice problems or a voice disorder history, (b) had any previous formal voice training or voice therapy, (c) were taking medication or had a medical condition that might affect normal voice function, (d) were intubated recently for any surgical intervention, or (e) had undergone surgery in the torso, head, and neck region in the last 18 months. These criteria, likely to influence jitter and shimmer values, were evaluated by questionnaire. Information on the participants' smoking habits and native language was also collected.

In addition, participants who were unable to phonate the sustained vowel /a/ for 5 s at different loudness levels after 10 min of training were excluded. Voice recordings were also not used if the mean GRBAS (grade, roughness, breathiness, asthenicity, strain) scale score (Hirano, 1981)—rated by two independent experts (one speech therapist and one phoniatrician) on the basis of three recorded phonations—was 1 or higher in any voice characteristic.

Included sample. A total of 57 participants—28 women 20–39 years of age (M = 28.8 years) and 29 men 20–39 years of age (M = 28.1 years)—were included for acoustic voice analysis. These included 9 (nondysphonic) smokers and 12 non-native speakers. Each participant provided 9 phonations, totaling 513 phonations for analysis. Of all 70 recruited participants, 12 participants with a mean GRBAS score of >1 in one voice characteristic and 1 participant unable to phonate for 5 s were excluded.

Procedures

Calibration. Conversion from uncalibrated voice signal amplitude (as measured by PRAAT; Boersma & Weenink, 2006) to calibrated SPL values was achieved using the comparison method (Winholtz & Titze, 1997). Prior to voice recordings, calibrated speech weighted noise (Wagener, Kühnel, & Kollmeier, 1999) was recorded with 10 cm distance to sound source at 50 dB, 65 dB, 85 dB, and 90 dB. The difference between known SPL from the calibrated signal and the measured uncalibrated amplitude values was calculated and was later used to compute calibrated SPL of the voice recordings.

Recording technique. Recordings were made according to European and American assessment guidelines (Dejonckere et al., 2001; Titze, 1995). Participants were recorded in a soundproof room (ambient noise approximately 20 dB) using a head-mounted off-axis positioned microphone (AKG Acoustics, C444) with 10-cm microphone–mouth distance and a portable DAT recorder (Sony,TCD-D8) at a sampling rate of 48000 Hz and 16-bit quantization. Later, each individual phonation was cut out and was anonymously labeled using the software program Audacity 1.2.4b (Mazzoni et al., 2005).

Training and experiment. During a training phase of up to 10 min, participants were asked to "sustain /a/ for 5 seconds at comfortable pitch and loudness." When able to do this, they were asked to "sustain /a/ for 5 seconds, as softly as possible" and then "as loudly as possible." These instructions were chosen to assess voice variance in uninfluenced voice function as done similarly in clinical practice. After training, when they were able to phonate at three recognizably different loudness levels (soft/medium/loud), participants were recorded phonating /a/ for 5 s at each level three times. The recording order was randomized in each case.

Data Analysis

Acoustic analysis. Acoustic analysis was conducted with PRAAT (Boersma & Weenink, 2006). To exclude the voice variability of the onset and offset phase, only the range from 0.5 s to 3.5 s after the voice onset was analyzed.

Jitter and shimmer measures can be subclassified into two main types: Absolute jitter and shimmer measures such as "perturbation factor" or "directional perturbation factor" are based on the difference between successive periods or amplitudes and are not normalized to the speaker's F_0 and SPL, respectively. Absolute jitter and shimmer measures have been shown to be affected by mean F_0 and SPL, respectively (Orlikoff & Baken, 1990). In contrast, F_0- and SPL-adjusted indexes are calculated as a ratio of mean perturbation to mean F_0 or amplitude, respectively. Thus, they are normalized to F_0 and SPL (Baken & Orlikoff, 2000a, 2000b). Therefore, a F_0-normalized jitter index, calculated as a percentage of F_0, was chosen. This parameter is called *jitter (local)* by PRAAT. Analogously, an SPL-adjusted index measuring shimmer in dB, called *shimmer (local, dB)* by PRAAT, was used. Because a different individual SPL and F_0 change was expected for "soft," "medium," and "loud" phonation between participants, and F_0 normally rises with increasing voice intensity (Gramming, 1988), SPL and F_0 were also measured.

Main outcome measures. The primary outcome measures were jitter (%) and shimmer (dB), with SPL and F_0 also measured as dependent variables. The most important independent variables for this study were phonation level (soft/medium/loud) and gender (f/m), although subject (1–57) and token (1–3 per loudness level) were also treated as independent variables.

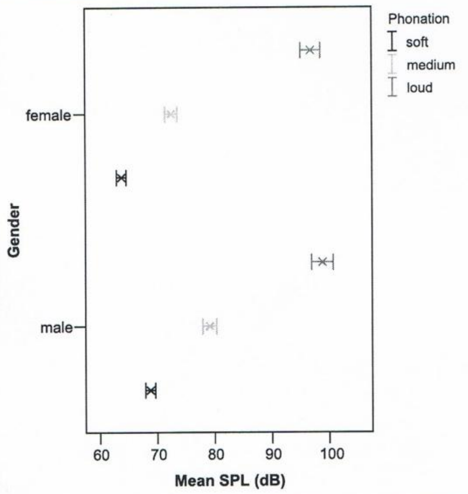
Statistical analysis. Mean values and 95% confidence intervals for SPL, F_0, jitter, and shimmer in soft, medium, and loud phonation were calculated for women and men. Additionally, the SPL distribution in gender was graphically displayed. The following inferential analysis was conducted for jitter and shimmer and for F_0 separately. First, since jitter, shimmer, and F_0 showed an approximately exponential relationship with SPL in the graphical display, a logarithmic transform was applied to each before further statistical analysis. Thereafter, the effects of four independent variables—phonation level (soft/medium/loud), gender (f/m), subject (1–57), and token (1–3 per loudness level)—were assessed with a four-way analysis of variance (ANOVA). To determine if effects due to soft, medium, and loud phonation were different in women and men, the interaction of phonation and gender was also assessed.

Results

SPL and F_0

Because the task was to produce soft, medium, and loud phonations, the mean SPLs produced in response to these three instructions are given in Figure1 and Table 2. Notably, there was an interaction between phonation level (soft/medium/loud) and gender, $F(2, 450) = 6.08$, $p < .01$ (see Table 3). Soft and medium phonations were significantly louder in men, whereas loud phonations were more similar between women and men (see Figure 1, Table 2).

Figure 1. Mean SPL and 95% confidence intervals (CIs) for soft, medium, and loud phonation in women and men.



From soft to medium phonation, there was little change in F_0 in both genders. However, F_0 rose significantly in men (+51 Hz) and women (+90 Hz) from medium to loud phonation (see Figure 2, Table 2). As expected for normal voices, gender had a significant effect on F_0, $F(1, 450) = 1625.90$, $p < .01$ (see Table 3).

Shimmer

Phonation level effects. There were highly significant differences in shimmer among phonation levels (soft/medium/loud), $F(2, 450) = 1054.81$, $p < .01$ (see Table 3). Over the whole SPL range, shimmer increased significantly with decreasing SPL (see Table 2); however, at SPLs below 80 dB (approximately the level of "comfortable" phonation), mean shimmer was considerably higher in both genders (see Figure 3).

Gender effects. Men had significantly smaller mean shimmer when asked to produce soft and medium phonations (see Table 2). The phonation level effects on shimmer were significantly different for women and men, $F(2, 450) = 21.62$, $p < .01$ (see Table 3). However, Figure 3, which illustrates actual SPL distribution and not phonation levels, shows no obvious gender effect at medium SPLs.

Jitter

Phonation level effects. As for shimmer, the phonation level had a highly significant effect on jitter, $F(2, 450) = 6.404$, $p < .01$ (see Table 3), which decreased with increasing SPL (see Table 2). Considerably higher jitter was observed below 75 dB in both genders (see Figure 4).

Gender effects. Men had significantly higher mean jitter in soft phonation (see Table 2). But in medium and loud phonation, jitter was similar between genders.

Discussion

This pragmatic study covers a wide SPL range in uninfluenced normal voice function in a typical clinical task and shows unequivocally that individual voice intensity has a considerable impact on both jitter and shimmer. Moreover, these effects are different for men and women. Though not reported here, similar results were measured in other shimmer and jitter types (such as "local shimmer", "rap," and "ddb") by PRAAT.

Response of SPL and F_0 to Soft, Medium, and Loud Phonation

Because SPL depends on distance, all measurements were made under standard conditions and were reported after comparison to reference values at 10 cm. Under these circumstances, mean SPL and F_0 in "medium"

Table 2. Mean and 95% confidence intervals for four instrumental parameters taken from soft, medium, and loud phonations.

	Soft phonation		Medium phonation		Loud phonation	
Variable	*F*	*M*	*F*	*M*	*F*	*M*
SPL (dB)	63.8 62.7–64.9	68.8 67.7–69.8	72.5 71.4–73.6	79.2 78.1–80.2	96.8 95.7–97.9	98.8 97.8–99.9
F_0 (Hz)	217 210–224	125 121–129	211 204–217	127 123–131	301 292–311	178 173–184
Shimmer (dB)	1.40 1.27–1.54	1.00 0.91–1.10	0.50 0.46–0.55	0.28 0.26–0.31	0.12 0.11–0.14	0.13 0.12–0.14
Jitter (%)	0.78 0.71–0.86	0.94 0.86–1.04	0.36 0.32–0.39	0.31 0.28–0.34	0.22 0.20–0.24	0.21 0.19–0.23

Note. Data are shown separately for women (F) and men (M). In all cases, there were significant differences between soft, medium, and loud phonations, with *p* values (between 10^{-85} and 10^{-214}) so low as to give essentially no probability of the results being explained by random variability. All *p* values have been rounded to three digits after the decimal point.

Table 3. ANOVA: Effects of gender (f/m), phonation level (soft/medium/loud), token (1–3), participant (1–57), and the Phonation × Gender interaction on SPL, F_0, jitter, and shimmer.

Factor	Dep. variable	*df*	*F*	*p*
Gender (f/m)				
	SPL	1	105.90	<.001
	F_0	1	1625.90	<.001
	Jitter (%)	1	0.00	.947
	Shimmer (dB)	1	52.94	<.001
Phonation level (soft/medium/loud)				
	SPL	2	1781.47	<.001
	F_0	2	307.77	<.001
	Jitter %	2	438.47	<.001
	Shimmer dB	2	1054.86	<.001
Token (1–3)				
	SPL	2	1.04	.354
	F_0	2	0.11	.892
	Jitter %	2	0.52	.592
	Shimmer dB	2	2.58	.077
Participant (1–57)				
	SPL	55	6.08	<.001
	F_0	55	10.17	<.001
	Jitter %	55	5.29	<.001
	Shimmer dB	55	6.85	<.001
Interaction: Phonation × Gender				
	SPL	2	9.39	<.001
	F_0	2	1.02	.36
	Jitter %	2	6.40	.002
	Shimmer dB	2	21.62	<.001

Note. *p* values have been rounded. ANOVA = analysis of variance; F_0 = fundamental frequency; Dep. = dependent.

phonation were consistent with "comfortable" phonation in young adults (Brown et al., 1996). The increase in F_0 from medium to loud phonation in both genders, and the higher SPL observed in men for soft and medium phonation, agree with previous observations (Gramming, 1988).

Response of Shimmer and Jitter to Soft, Medium, and Loud Phonation

Our jitter and shimmer measurements at medium loudness (see Table 2) were in broad agreement with the

Figure 2. SPL and fundamental frequency (F_0) distribution in women and men over the complete range of phonatory levels.

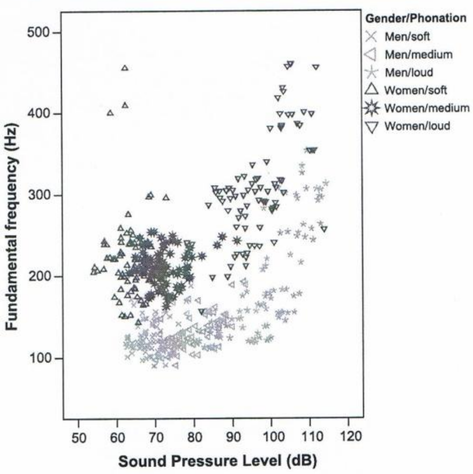


Figure 3. SPL and shimmer relationship. Shimmer increases intensely in phonations below 80 dB in both genders.

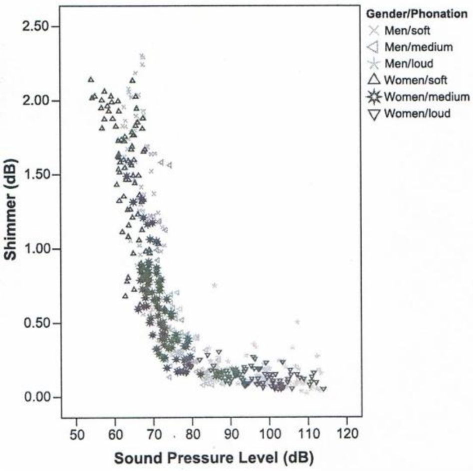
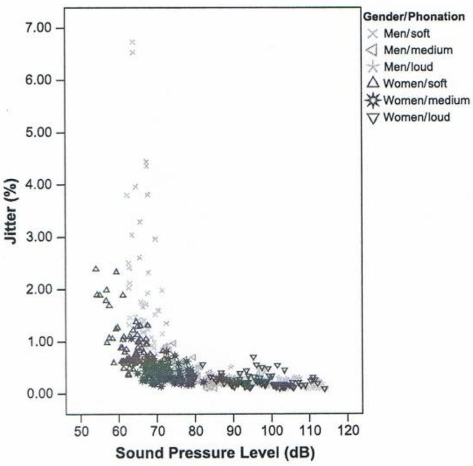


Figure 4. SPL and jitter relationship. Jitter increases in phonations below 75 dB in both genders.



data from other reports at comfortable loudness and pitch (see Table 1). However, we have shown that shimmer and jitter depend on phonation level to an extent that has not previously been reported. The decrease in jitter and shimmer with increasing SPL was also observed in supplemented phonetograms of healthy adults (Pabon, 1991; Pabon & Plomp, 1988), in healthy and dysphonic adults at unspecified SPL (Dejonckere, 1998), and in women at predefined SPL (Gelfer, 1995). The mean shimmer and jitter in "medium" and "loud" male phonation in this study were comparable to measurements in a defined SPL range at "moderate" and "loud" level (Orlikoff & Kahane, 1991). However, our results show a dramatic quasiexponential increase in shimmer and jitter below 75–80 dB (see Figures 3 and 4) rather than the linear relationship as reported previously (Orlikoff & Kahane, 1991). The contradiction probably reflects the wide observed SPL range of 63.8–98.8 dB in our work, as compared with 72.8–94.3 dB (converted to 10 cm distance) reported by Orlikoff and Kahane (1991).

Acoustic analysis at different voice SPL might also partially explain contradictory jitter and shimmer values reported for women and men in previous studies. It is noteworthy that our participants did not show differences based on gender, as shown very clearly in Figure 3. This was unexpected and opens up the perspective that in the future, normative values should be established for a range of SPLs.

Gender Effects on Jitter and Shimmer

Gender effects on jitter and shimmer are clearly smaller than the extremely large effects due to phonation level (see Figures 3 and 4). Ostensibly, the lower shimmer in men in soft and medium phonation (see Table 3) agrees with findings "at comfortable loudness and pitch" (Sorensen & Horii, 1983; Jafari et al., 1993; see Table 1). However, it is noteworthy that Figure 3 does not support the idea that men have lower shimmer in "medium" SPLs. Indeed, when using actual SPL (and not the phonation levels of soft, medium, and loud) as the predictive factor, then the gender effect disappears at medium SPL ranges.

In our pragmatic study, we investigated the performance of women and men in a typical clinical voice assessment task, when producing "medium" or "normal" voice loudness. Although there is measurably less shimmer in men when phonating at medium loudness, our data suggest that this is simply because men tend to phonate louder. As clearly shown in Figure 3, for men and women phonating at the same SPL, there would be little or no difference at medium voice loudness, which was also observed by other authors (Deem et al., 1989; Ludlow et al., 1987). This effect probably explains some of the contradictory results reported previously, although their methods were not reported in enough detail to appraise this.

Gender effects might be also associated to a F_0 change alone, especially in men (Orlikoff & Baken, 1990). Because

F_0 did not change much between soft and medium phonation, the observed effects are most likely due to the observed SPL change.

Explanation Model

Greater jitter and shimmer below 80 dB might reflect a physiological laryngeal tension change in normal voices associated with low SPL. In lower intensities and frequencies, a higher glottal open quotient, associated with lower intrinsic vocal fold muscle tension and smaller vocal fold vibration amplitudes, has previously been described (Hodge et al., 2001; Sulter, Schutte, & Miller, 1996). Especially low intrinsic vocal fold muscle tension might result in greater mucosa cover variability, generating higher voice perturbation in soft phonation. So jitter and shimmer might track subtle tension related to voice changes. This would have a significant clinical value in diagnosing subtle voice disorders such as muscle tension dysphonia, which are difficult to diagnose by other means including videolaryngostroboscopy (Schneider, Wendler, & Seidner, 2002).

Clinical Implications

Reliability and sensitivity. One key concern is that in both genders, shimmer and jitter increased dramatically when the participants phonated below 80 and 75 dB, respectively. When adults are asked to phonate at a "comfortable level" as done in clinical assessments, they are likely to produce less than 75 dB (Brown et al., 1996). In our study, women and men produced voice below 75 dB and 80 dB, respectively, in "medium" phonation. So, in normal voice assessment, it is likely that healthy young adults phonate at critically low SPL. Reported unsatisfactory reliability and sensitivity (Carding et al., 2004; Zyski et al., 1984) might be partially explained by the fact that jitter and shimmer are higher and vary substantially more at this loudness level. Furthermore, women are at risk of appearing more pathological because they phonate more quietly in medium phonation. This suggests that normative thresholds are not simply applicable in practice and transferable between genders without adequate SPL control.

Improving measurement accuracy. The effect of SPL and gender on jitter and shimmer could be minimized by requesting phonations above 80 dB (as recorded at 10 cm). This equates to adult voice loudness when speaking to a listener 9 m away (Healey, Jones, & Berky, 1997). An SPL meter with visual feedback would easily facilitate SPL control in voice assessments. However, future work is required to investigate how voluntarily controlling SPL affects jitter and shimmer measurements.

Although it can be achieved in functional dysphonia (Gramming, 1988) and polyps (Pabon & Plomp, 1988), phonation above 80 dB may be impossible in some pathologies such as laryngeal palsy. Therefore, whether voice perturbation analysis can usefully be applied in severely disordered voices is open to question. This criticism has also been raised from a different angle, as severely hoarse voices can be too aperiodic to provide meaningful jitter and shimmer measurements (Titze, 1995). Furthermore, results may not be simply transferable to groups in which modestly different jitter and shimmer can be presumed, such as older adults (Wilkox & Horii, 1980) or mildly to moderately dysphonic patients (Schoentgen, 1982). These groups would have to be differentially examined in future investigations.

Conclusion

This pragmatic study provides baseline data for the effects of voice loudness and gender on acoustic voice parameters in healthy young adults. Both jitter and shimmer increase dramatically as SPL decreases, especially between "medium" to "soft" voice loudness. Below 80 dB, small variations in SPL lead to large changes in jitter and shimmer. These low SPL intensities are likely associated with small intrinsic vocal fold tension and vibration amplitudes, potentially causing more mucosa movement variability, and resulting in higher perturbation.

At any given SPL, the levels of jitter and shimmer appeared comparable between men and women. However, in a real clinical situation, when asked for "comfortable" voice loudness, women tend to respond with a softer phonation and so may appear more pathologic. This effect was particularly evident in shimmer measurements. Therefore, normative thresholds are not simply applicable and transferable between genders.

SPL control potentially could be used to enhance the reliability of voice assessment—for example, by requesting phonations above 80 dB. However, it is unclear how jitter and shimmer respond to SPL control or how SPL affects perturbation in disordered voices or other age groups. These questions have to be addressed in further research to ensure jitter and shimmer measurement reliability.

References

Askenfelt, A., & Hammarberg, B. (1986). Speech waveform perturbation analysis: A perceptual acoustical comparison of seven measures. *Journal of Speech and Hearing Research, 29,* 50–64.

Baken, R., & Orlikoff, R. F. (2000a). Amplitude perturbation (shimmer). In R. Baken & R. F. Orlikoff (Eds.), *Clinical measurements of speech and voice* (pp. 130–137). Clifton Park, NY: Thomson Delmar Learning.

Baken, R., & Orlikoff, R. F. (2000b). Frequency perturbation (jitter). In *Clinical measurements of speech and voice* (pp. 190–213). Clifton Park, NY: Thomson Delmar Learning.

Boersma, P., & Weenink, D. (2006). Praat (Version 4.4.16) [Voice analysis computer software]. Amsterdam, the Netherlands: University of Amsterdam. Retrieved January 07, 2006, from http://www.praat.org/.

Brown, W. J., Morris, R., & Murry, T. (1996). Comfortable effort level revisited. *Journal of Voice, 10,* 299–305.

Carding, P. N., Steen, I. N., Webb, A., Mackenzie, K., Deary, I. J., & Wilson, J. A. (2004). The reliability and sensitivity to change of acoustic measures of voice quality. *Clinical Otolaryngology, 29,* 538–544.

Deem, J., Manning, W., Knack, J., & Matesich, J. (1989). The automatic extraction of pitch perturbation using microcomputers: Some methodological considerations. *Journal of Speech and Hearing Research, 32,* 689–697.

Dejonckere, P. (1998). Effect of louder voicing on acoustical measurements in dysphonic patients. *Logopedics, Phoniatrics, Vocology, 23,* 79–84.

Dejonckere, P., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchmann, L., Friedrich, G., et al. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Otorhinolarygology, 258,* 77–82.

Dejonckere, P., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchmann, L., & Millet, B. (1996). Differentiated perceptual evaluation of pathological voice quality: Reliability and correlations with acoustic measurements. *Revue de laryngologie, d'otologie et de rhinologie, 117*(3), 219–224.

Eksteen, E., Rieger, J., Nesbitt, M., & Seikaly, H. (2003). Comparison of voice characteristics following three different methods of treatment for laryngeal cancer. *Journal of Otolaryngology, 32*(4), 250–253.

Ferrand, C. (1995). Effects of practice with and without knowledge of results on jitter and shimmer levels in normally speaking women. *Journal of Voice, 9*(4), 419–423.

Gelfer, M. P. (1995). Fundamental frequency, intensity, and vowel selection: Effects on measures of phonatory stability. *Journal of Speech and Hearing Research, 38,* 1189–1198.

Gramming, P. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice, 2,* 118–126.

Hanson, D. G., Jiang, J. J., Chen, J., & Pauloski, B. R. (1997). Acoustic measurement of change in voice quality with treatment for chronic posterior laryngitis. *Annals of Otology, Rhinology and Laryngology, 106*(4), 279–285.

Healey, E., Jones, R., & Berky, R. (1997). Effects of perceived listeners on speakers' vocal intensity. *Journal of Voice, 11*(1), 67–73.

Hirano, M. (1981). *Clinical examination of voice* (Vol. 5). New York: Springer-Verlag.

Hodge, F., Colton, R., & Kelley, R. (2001). Vocal intensity characteristics in normal and elderly speakers. *Journal of Voice, 15*(4), 503–511.

Horii, Y. (1980). Vocal shimmer in sustained phonation. *Journal of Speech and Hearing Research, 23,* 202–209.

Horii, Y. (1982). Jitter and shimmer differences among sustained vowel phonations. *Journal of Speech and Hearing Research, 25,* 12–14.

Jafari, M., Till, J., Truesdell, L., & Law-Till, C. (1993). Time-shift, trial and gender effects on vocal perturbation measurements. *Journal of Voice, 7,* 326–336.

Kay Elemetrics Corp. (1999). Multi-Dimensional Voice Program (MDVP) Model 5105: Instruction manual. (Version 2.0) [Acoustic Analysis]. Lincoln Park, NJ: Kay Elemetrics Corporation.

Ludlow, C., Bassich, C., Connor, N., Connor, N., Coulter, D., & Lee, Y. (1987). The validity of using jitter and shimmer to detect laryngeal pathology. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration: Vocal fold physiology series* (pp. 492–508). Boston: College-Hill.

Mazzoni, D., Brubeck, M., Crook, J., Johnson, V., Meyer, M., Habermann, J., et al. (2005). Audacity (Version 1.2.4b) [Audio Editor and Recorder]. Boston, MA: Free Software Foundation.

Orlikoff, R. F., & Baken, R. (1990). Consideration of the relationship between the fundamental frequency of phonation and vocal Jitter. *Folia Phoniatrica, 42,* 31–40.

Orlikoff, R. F., & Kahane, J. C. (1991). Influence of mean sound pressure level on jitter and shimmer measures. *Journal of Voice, 5*(2), 113–119.

Pabon, J. P. (1991). Objective acoustic voice-quality parameters in the computer Phonetogram. *Journal of Voice, 5*(3), 203–216.

Pabon, J. P., & Plomp, R. (1988). Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *Journal of Speech and Hearing Research, 31,* 710–722.

Rovirosa, A., Martinez-Celdran, E., Ortega, A., Ascaso, C., Abellana, R., Velasco, M., et al. (2000). Acoustic analysis after radiotherapy in T1 vocal cord carcinoma: A new approach to the analysis of voice quality. *International Journal of Radiation Oncology, Biology, Physics, 47,* 73–79.

Roy, N., Weinrich, B., Gray, S., Tanner, K., Toledo, S., Dove, H., et al. (2002). Voice amplification versus vocal hygiene instruction for teachers with voice disorders: A treatment outcomes study. *Journal of Speech Language & Hearing Research, 45,* 625–638.

Schneider, B., Wendler, J., & Seidner, W. (2002). The relevance of stroboscopy in functional dysphonias. *Folia Phoniatrica, 54,* 44–54.

Schoentgen, J. (1982). Qualitative evaluation of the discriminating performance of acoustic features in detecting laryngeal pathology. *Journal of Speech and Hearing Research, 13,* 489–502.

Sorensen, D., & Horii, Y. (1983). Frequency and amplitude perturbation in the voices of female speakers. *Journal of Communication Disorders, 16,* 57–61.

Sulter, A., Schutte, H., & Miller, D. (1996). Standardized laryngeal videostroboscopic rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age. *Journal of Voice, 10*(2), 175–189.

Titze, I. R. (1995). Workshop on acoustic voice analysis: Summary statement. In D. Wong (Ed.), *Proceedings of the Workshop on Acoustic Voice Analysis,* Denver, CO (February 1994). Iowa City, IA: University of Iowa.

Wagener, F., Kühnel, V., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztestes in deutscher Sprache. I: Design des Oldenburger Satztestes (German) [Development and evaluation of a sentence-based hearing test in German language. I: Design of the Oldenburger Satztest]. *Zeitschrift für Audiologie, 38,* 4–15.

Wilkox, K., & Horii, Y. (1980). Age and changes in vocal jitter. *Journal of Gerontology, 35*(2), 194–198.

Winholtz, W., & Titze, I. (1997). Conversion of a head-mounted microphone signal into calibrated SPL units. *Journal of Voice, 11*(4), 417–421.

Zyski, B., Bull, G., McDonald, W., & Johns, M. (1984). Perturbation analysis of normal and pathological larynges. *Folia Phoniatrica, 36,* 190–198.

Received November 13, 2006

Revision received June 5, 2007

Accepted November 29, 2007

DOI: 10.1044/1092-4388(2008/06-0208)

Contact author: Meike Brockmann, Head of Speech Pathology Section, Clinic of Otorhinolaryngology, Head and Neck Surgery, University Hospital Zurich, Frauenklinikstrasse 24, 8091 Zürich, Switzerland. E-mail: meike.brockmann@usz.ch.

Claudio Storck is now with the Department of Phoniatrics, Clinic of Otorhinolaryngology, Head and Neck Surgery, University Hospital Basel, Switzerland.

Appendix C: Ethical Approval Information

Ethical approval was granted for all studies by the local ethical committee responsible for the University Hospital Zurich ("Kantonale Ethikkommission Zürich, KEK"). The KEK Zurich assigned reference numbers to each study as listed below:

Chapters 4 and 5: approval study reference number 612

Chapter 6: approval of amendment to study reference number 612

Chapter 8: approval study reference number 832

Appendix D: Copy of ethical approval studies chapters 4 and 5 with translation

Formular für die
Beschlussmitteilung der Ethikkommission

Die SPUK ZH Spezialfächer hat an ihrer Sitzung vom 09.12.05 (in der Zusammensetzung, wie sie nachstehend wiedergegeben ist) das folgende Forschungsprojekt eingehend begutachtet.

Titel des Forschungsprojektes Ref.Nr. EK: **Nr. 612**

Untersuchung des Einflusses von Stimmlautstärke und Stimmtonhöhe auf die diagnostisch angewandten objektiven akustischen stimmlichen Parameter Jitter und Shimmer bei stimmgesunden erwachsenen Personen

Prüfer/in (verantwortliche Studienleiter/in am Versuchsstandort)

Name, Vorname, Titel:	Storck, Claudio, Dr. med.
Funktion:	Oberarzt, Phoniater
Adresse:	Klinik für Ohren-, Nasen-, Hals- und Gesichtschirurgie, Abt. für Phoniatrie, USZ, Frauenklinikstrasse 24, 8091 Zürich

Die Ethikkommission stützt ihre Beurteilung auf die Unterlagen, wie sie dem beiliegenden „Basisformular zur Einreichung eines biomedizinischen Forschungsprojektes" vom 21.12.05 (Datum) beigefügt sind.

☐ normales Verfahren ☐ vereinfachtes Verfahren ☒ Nachbegutachtung

Die Ethikkommission kommt zu folgendem **Beschluss**:

☒ A **positiv**
☐ B **positiv mit Empfehlungen** (siehe Seite 2ff)
☐ C **mit Auflagen** (siehe Seite 2ff)
Nachbegutachtung durch Ethikkommission notwendig ☐
schriftliche Mitteilung an Ethikkommission ausreichend ☐
☐ D **negativ (mit Begründung und Erläuterung für die Neubeurteilung)** (siehe Seite 2ff)
☐ E **Nicht-Eintreten (mit Begründung)** (siehe Seite 2ff)

Der Beschluss gilt auch für die im "Basisformular" gemeldeten weiteren Prüfer/innen im Zuständigkeitsbereich der Ethikkommission.

Empfehlungen

(erweiterbar)

Translation ethical approval studies chapters 4 and 5

Form for notification of **ethical review committee decisions**

The SPUK ZH specialty panel has examined the following project in its meeting from 09.12.2005 (participants are listed below).

Title of project Reference **No. 612**

Examination of the influence of voice loudness and pitch on the diagnostic objective acoustic voice analysis parameters jitter and shimmer in vocally healthy adults

Examiner (responsible research director at the institute)

Surname, name, title:	Storck, Claudio, Dr. med.
Capacity:	Senior Physician, Phoniatrician
Address:	Clinic for Otorhinolaryngology, Head and Neck Surgery; Dept. for Phoniatrics, USZ (University Hospital Zurich), Frauenklinikstrasse 24, 8091 Zurich

The decision of the ethical review board is based on the submitted paperwork as enclosed with the "Basic Form for Submission of a Biomedical Research Project" from 20.02.2009.

□ standard procedure □ simplified procedure X second assessment

Ethical review board **decision**:

X **A positive**

□ **B positive with recommendation** (see page 2 ff.)

□ **C with requirements** (see page 2 ff.)

Second assessment through ethical review board necessary □

Written report to ethical review board sufficient □

□ **D negative (with explanatory statement for resubmission)** (see page 2 ff.)

□ **E rejection** (see page 2 ff.)

The decision applies to all in the "Basic Form" named examiners in the ethical review board field of responsibility.

Recommendations

Appendix E: Copy of ethical approval study chapter 6 with translation

Kantonale Ethikkommission

Spezialisierte Unterkommission für Spezialfächer

Prof. Dr. med. J.A. Fischer
Präsident

Kantonale Ethikkommission
SPUK für Spezialfächer
UniversitätsSpital Zürich
Sonneggstrasse 12
CH-80091Zürich
Tel. 044 255 95 21
Fax 044 255 95 22
E-Mail helene.viera@kaz.zh.ch
Internet: http://www.ethikkommission.ch

Herrn Prof. Dr. med. R. Probst
Frau M. Brockmann
Klinik für Ohren- Nasen-, Hals- und
Gesichtschirurgie
Universitätsspital Zürich
Frauenklinikstrasse 24
8091 Zürich

Zürich, 26. Oktober 2007

Antrag Nr. 612 - Amendment Nr. 3 vom 10. Oktober 2007
Untersuchung des Einflusses von Stimmlautstärke und Stimmtonhöhe auf die diagnostisch angewandten objektiven akustischen stimmlichen Parameter Jitter und Shimmer bei stimmgesunden erwachsenen Personen

Sehr geehrter Herr Professor Probst

Besten Dank für Ihren Brief vom 10. Oktober 2007 mit dem Amendment Nr. 3 und den folgenden Unterlagen:

- Zusammenfassung der bisherigen Studienergebnisse zur oben genannten Studie
- Probandeninformation und –einverständniserklärung vom 10. Oktober 2007
- Prüfprotokoll vom 10. Oktober 2007

Das Amendment wurde vom Präsidenten geprüft. Da es sich nur um Änderungen handelt, welche die Sicherheitslage für die Patienten nicht beeinträchtigen, muss das Amendment nicht erneut an einer Kommissionssitzung traktandiert werden und ist damit bis 31.03.2008 bewilligt.

Die Gebühr für Amendments wird von der Kantonalen Ethikkommission erhoben.

Mit freundlichen Grüssen

Prof. Dr. med. Jan A. Fischer
Präsident SPUK für Spezialfächer

Kopie an:
- Kantonale Ethikkommission (KEK)

Translation ethical approval study chapter 6

Prof. Dr. med. R. Probst
Ms. M. Brockmann
Clinic for Otorhinolaryngology, Head and Neck Surgery
University Hospital Zurich
Frauenklinikstrasse 24
8091 Zurich

Zurich, 26th October 2007

Proposal No. 612- Amendment No. 3 from 10. October 2007
Examination of the influence of voice loudness and pitch on the diagnostic objective acoustic analysis parameters jitter and shimmer in vocally healthy adults

Dear Prof. Probst,
Thank you for your letter from 10. October 2007 with amendment No. 3 and the following documents:

- Summary of previous study results
- Written information for study participants
- Test protocol from 10. October 2007

The amendment was examined by the president of the ethical review board. You applied for modifications to the test protocol, which do not affect patient safety. Therefore the amendment does not have to be presented at a full ethical review board meeting and is approved until 31.03.2008.
The obligatory fees for amendments will be charged by the ethical review board.
Kind regards

Prof. Dr. med. Jan A. Fischer
President SPUK specialty panel

Copy to: Cantonal Review Board (Kantonale Ethikkommission KEK)

Appendix F: Copy of ethical approval study chapter 8 with translation

Formular für die
Beschlussmitteilung der Ethikkommission

Die SPUK ZH Spezialfächer hat an ihrer Sitzung vom 12.12.08 (in der Zusammensetzung, wie sie nachstehend wiedergegeben ist) das folgende Forschungsprojekt eingehend begutachtet.

Titel des Forschungsprojektes Ref.Nr. EK: **Nr. 832**

Untersuchung der Sensitivität der objektiven akustischen Stimmanalyse in der Diagnostik von diskreten stimmlichen Symptomen

Prüfer/in (verantwortliche Studienleiter/in am Versuchsstandort)

Name, Vorname, Titel:	Probst, Rudolf, Prof. Dr. med.
Funktion:	Klinikdirektor
Adresse:	ORL-Klinik, USZ, Frauenklinikstrasse 24, 8091 Zürich

Die Ethikkommission stützt ihre Beurteilung auf die Unterlagen, wie sie dem beiliegenden „Basisformular zur Einreichung eines biomedizinischen Forschungsprojektes" vom 20.02.09 (Datum) beigefügt sind.

☐ normales Verfahren ☐ vereinfachtes Verfahren ☒ Nachbegutachtung

Die Ethikkommission kommt zu folgendem **Beschluss**:

☒ **A positiv**
☐ **B positiv mit Empfehlungen** (siehe Seite 2ff)
☐ **C mit Auflagen** (siehe Seite 2ff)
Nachbegutachtung durch Ethikkommission notwendig ☐
schriftliche Mitteilung an Ethikkommission ausreichend ☐
☐ **D negativ (mit Begründung und Erläuterung für die Neubeurteilung)** (siehe Seite 2ff)
☐ **E Nicht-Eintreten (mit Begründung)** (siehe Seite 2ff)

Der Beschluss gilt auch für die im "Basisformular" gemeldeten weiteren Prüfer/innen im Zuständigkeitsbereich der Ethikkommission.

Empfehlungen

(erweiterbar)

Translation ethical approval study chapter 8

Form for notification of **ethical review committee decisions**

The SPUK ZH specialty panel has examined the following project in its meeting from 12.12.2008 (participants are listed below).

Title of project Reference **No. 832**

Examination of the sensitivity of objective acoustic voice analysis parameters in the diagnostics of discrete voice symptoms

Examiner (responsible research director at the institute)

Surname, name, title:	Probst, Rudolf, Prof. Dr. med.
Capacity:	Head of Clinic
Address:	Clinic for Otorhinolaryngology, Head and Neck Surgery; USZ (Univ. Hospital Zurich), Frauenklinikstrasse 24, 8091 Zurich

The decision of the ethical review board is based on the submitted paperwork as enclosed with the "Basic Form for Submission of a Biomedical Research Project" from 20.02.2009.

□ standard procedure □ simplified procedure X second assessment

Ethical review board **decision**:

X **A positive**

□ **B positive with recommendation** (see page 2 ff.)

□ **C with requirements** (see page 2 ff.)

Second assessment through ethical review board necessary □

Written report to ethical review board sufficient □

□ **D negative (with explanatory statement for resubmission)** (see page 2 ff.)

□ **E rejection** (see page 2 ff.)

The decision applies to all in the "Basic Form" named examiners in the ethical review board field of responsibility.

Recommendations

Appendix G: Participant questionnaire for studies chapters 4-8

(Translated from German)

Dear Mr / Ms ____________________________ *(Name of participant)*

Please read the following questions carefully and answer them. If you have further questions concerning the questionnaire, please do not hesitate to ask the examiner.

	Yes	No
Have you taken individual singing or voice training lessons? Please specify. For how many months/years? ________________________	☐	☐
Have you ever had a voice disorder or have you received treatment because of voice problems?	☐	☐
Do you have a hearing disorder?	☐	☐
Do you take any medication on a regular basis? Please specify: ____	☐	☐
Do you have a neurologic or psychiatric disease?	☐	☐
Have you had a tumour or an operation in the torso, head or neck in the last 18 months?	☐	☐
Have you been intubated for any surgery in the last 18 months?	☐	☐
Have you had an upper airway infection in the last month?	☐	☐
Do you suffer from reflux or heartburn?	☐	☐
Do you presently suffer from an allergy (for example, hay fever)?	☐	☐
Do you smoke?	☐	☐
Is Swiss German your native language or first language? Which language do you mainly speak at home? _______________________________________	☐	☐

____

(Date, signature participant)

____

(Date, signature examiner)

(This page was kept separately)

Date: ______________ Participant number: ______________

Name: ________________________________ Gender: ___________

Date of birth: __________________________ Age: ______________

GRBAS Score:____

GRBAS Score:____

Comments:____

____

____

____

WAV-File signature: _____________________________

Appendix H: Oral instruction for judges (studies chapters 7 and 8)

(Translated from German)

Dear colleagues,

Thank you for participating in this study. The focus of this research is to examine how the human ear perceives „irregularity“ in the sound of healthy voices and how good this compares to irregularity measurements by computer analysis.

As you know healthy voices can sound very different from each other. Your task in this research work is to rank voice recordings of healthy adults according to „irregularity“ as you hear it in the voice sound.

There will be 4 groups to judge with 10 voice recordings each. Two groups contain healthy female voices and the two other groups contain 10 healthy male voices. All recorded participants say the vowel /a/ for around 5 seconds at comfortable voice intensity.

You will be able to play the voice samples with computer software named „Newcastle Audio Ranking“, which I will explain later. The most important aspect of your task is to listen carefully to the voice samples in one group and to create a rank order of all recordings according to perceptual „irregularity“.

Furthermore you will be asked to repeat exactly the same ranking task with the same voice recordings in one week. This will give us an idea about the variability of the perception of “irregularity” in normal voices.

Appendix I: User manual Software Newcastle Audio Ranking (studies chapters 7 and 8)

(Translated from German)

Dear colleague,

Thank you for participating in this study. This manual explains how the software, which you will use for the ranking task, works. If there are questions to the program, please do not hesitate to ask Meike Brockmann.

Staring NeAR and choosing a subset for the ranking session

Please start the NeAR program with a double click (left mouse button) on the *Newcastle Audio Ranking* folder.

Please use the *Choose Folder* button to pick out one folder.

On the memory stick you will find 4 folders numbered 1 to 4. Each folder contains all recordings of one subset. Please pick one folder.

In the box underneath the *Choose Folder* button it should indicate the folder name and the number of files in it to rank.

Choosing a name to identify the ranking session

To make the analysis easier, all the results are saved in one file. This helps us to identify rating sessions.

Please type in “1” for the ranking of recording subset 1 *(folder 1), “2” for recording subset two, “3” for recording subset three and “4” for subset four.*

You could use the same name twice, but you will receive a warning. Please do not rank the same subset twice or use the same name twice.

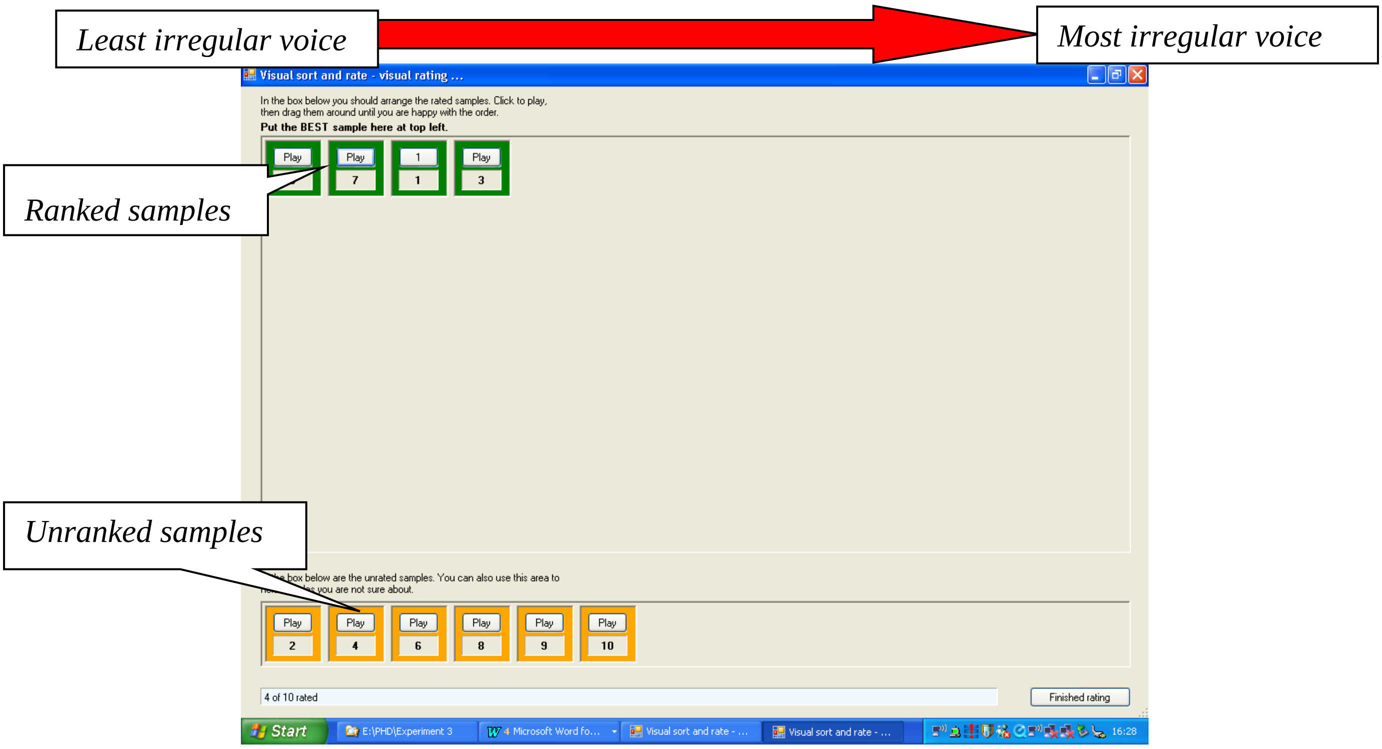
Please note that the date and time of the session will also be saved.

How to start a ranking session

Click on the *Start* button, and the ranking window appears. The ranking process works in the following steps:

- The *unranked samples* appear at the bottom of the screen in orange.
- Click the *Play* button on any sample to hear it.
- Then, drag it up to the top of the screen using the left mouse button. It will turn blue as you drag it, then green when you drop to indicate that it has been ranked.

- Repeat this for the second sample. *Drop it to the left of the first sample if you think the voice is less irregular*. Drop it to the right of the first sample, if you think the voice is more irregular (please see illustration). And so on.
- You can play and move the samples as often as you like, until you are happy with the order.
- If you are not sure about a sample, drag it back to the bottom of the screen where it re-joins the unranked samples. Please note down if you feel unable to rate a specific sample and give the reasons.



End the ranking session

To finish the ranking session please click the *Finished rating* button.

The results are saved automatically, and you can start another ranking session.

Further questions..........

If there are further questions please ask:

Meike Brockmann MSc, Head of Speech Pathology, ORL USZ

meike.brockmann@usz.ch

Tel: 044/ 35 55830

Appendix J: Visual Analogue Scale subjective voice assessment (study chapter 8)

(Translated from German)

Examiner: _____ Participant Nr._____

How good is your voice?

1. |____|

Very bad very good

2. |____|

Very bad very good

3. |____|

Very bad very good

4. |____|

Very bad very good