Doctoral Thesis

Selected topics in laryngeal, perceptual and acoustic assessments of human voice: Videokymographic evaluations of vocal folds and investigations of teachers' voices

Ketaki Vasant Phadke, M.Sc.

Department of Biophysics
Faculty of Science
Palacký University Olomouc, Czech Republic

Supervisor: RNDr. Jan G. Švec, Ph.D. et Ph.D.

2018

Ketaki Vasant Phadke

Selected topics in laryngeal, perceptual and acoustic assessments of human voice: Videokymographic evaluations of vocal folds and investigations of teachers' voices

Vybraná témata k hodnocení lidského hlasu pomocí vyšetření hrtanu, percepčních a akustických metod: Videokymografické hodnocení hlasivek a výzkum hlasu učitelů

Doctoral Thesis, Palacký University Olomouc, Czech Republic
– With an abstract in Czech.

Inspired by ‘The Law of Attraction’

“Faith is taking the first step even when you can’t see the whole staircase”

-Dr. Martin Luther King, Jr.

What We Think, We Create
What We Feel, We Attract
What We Imagine, We Become

-Buddha?

Table of Contents

ACKNOWLEDGEMENT

"Feeling gratitude and not expressing it is like wrapping a present and not giving it"

-WILLIAM ARTHUR WARD

I express my humble gratitude to several incredible people who played a vital role during my doctoral study and in helping me to complete my thesis.

I extend my deepest gratitude to ***RNDr. Jan Švec*** *for accepting me as his doctoral student and supervising me during my doctoral studies. I thank him for his persistent help in bringing up this thesis to its current level. Dr. Švec's continual support, teaching and encouragement have helped me to grow and progress well as a researcher. Devoting his time and helping me to understand every concept, with extreme patience is worth being grateful for. His time dedicated in developing my academic, research and writing skills is invaluable. I also thank him for supporting me with funds and providing me opportunities to present my research results at various international conferences and forums.*

I express my sincere gratitude to ***Prof. RNDr. Petr Ilík*** *(Head of the Department of Biophysics), for giving me the opportunity to pursue my doctoral studies at the Department of Biophysics, and supporting me with doctoral scholarship and funds during my research stay at Helsinki, Finland. I also thank all the staff and faculty members of Department of Biophysics for their support.*

I am indebted to **Dr. Ahmed Geneid** *(Head of the Department of Otorhinolaryngology and Phoniatrics, University of Helsinki, Finland) and* **Dr. Anne-Marie Laukkanen** *(University of Tampere, Finland) for involving me in their projects and supporting me with my studies during my three months research stay at Helsinki, Finland.*

I take this opportunity to thank my colleague, **Dr. Pravin Kumar Subbaraj***, for his help during our publication and making me understand some physics/signal processing concepts through his guidance and teaching.*

I express my thanks to **Irma***, for her continuous help, support and motivation during these four years. I also thank all my friends in Czech Republic, India and around the world for their encouragement and support*

A four walled edifice becomes a place of comfort living when you have the best family. It is the place where life begins and love never ends. I thank my wonderful ***Ayee***, ***Baba*** *and* ***Tai*** *for their love and support. Their love and presence in my life is paramount and unparalleled to anything else. In addition, a special thanks to my best friend and companion,* **Dr. Abhishek Mani Tripathi** *for his love, persistent support, encouragement and guidance.*

This study was supported by:

- *The Technology Agency of the Czech Republic (**TAČR**) project no. TA04010877.*
- *The Czech Science Foundation (Grantová Agentura České Republiky -* ***GAČR****) project no. GA16-01246S.*

My three months research stay was supported by the funds from **Erasmus Plus** *program (grant no. 2017–2018/164, Palacký University Olomouc, Czech Republic).*

LIST OF PUBLICATIONS IN THIS THESIS

The thesis is based on the following publications. The manuscripts are enclosed at the end of the thesis, which are referred in the text by the corresponding roman numerals:

I. Phadke, K.V., Vydrová, J., Domagalská, R. and Švec, J.G., 2017. Evaluation of Clinical Value of Videokymography for Diagnosis and Treatment of Voice Disorders. *European Archives of Oto-Rhino-Laryngology*, *274*(11), pp.3941-3949. https://doi.org/10.1007/s00405-017-4726-1

- Received **2nd place in the Student Scientific competition for the Dean's Award-2018**) at Přírodovědecká fakulta UP, 17. listopadu 12, Olomouc, in the section Biologie a Ecologie under doctoral division.

- Received the **2nd Annual Gisele Oliveira Award** for poster presentation entitled *"Contribution of High speed Videokymography to diagnosis of voice disorders-A Clinical study"* at 'Voice Foundation Annual Symposium: Care of the Professional Voice-2017, Philadelphia, USA.

II. Kumar, S.P., Phadke, K.V., Vydrová, J., Novozámský, A., Zita, A., Zitová, B. and Švec, J.G., 2018. Visual and Automatic Evaluation of Vocal Fold Mucosal Waves Through Sharpness of Lateral Peaks in High-Speed Videokymographic Images. *Journal of Voice* (*In press, early online*). https://doi.org/10.1016/j.jvoice.2018.08.022

- This research paper won the "**Hamdan International Presentation Award**" and "**Sataloff New Investigator award**" at the Voice foundation: Annual Symposium: Care of the professional voice-2018, Philadelphia, USA. Award was received by the first and presenting author: Dr. Pravin K. Subbaraj.

III. Phadke, K.V., Laukkanen, A.M., Ilomäki, I., Kankare, E., Geneid, A. and Švec, J.G., 2018. Cepstral and Perceptual Investigations in Female Teachers with Functionally Healthy Voice. *Journal of Voice* (*In press, early online*). https://doi.org/10.1016/j.jvoice.2018.09.010

IV. Phadke, K.V., Abo-Hasseba, A., Švec, J.G. and Geneid, A., 2018. Influence of Noise Resulting From the Location and Conditions of Classrooms and Schools in Upper Egypt on Teachers' Voices. *Journal of Voice* (*In press, early online*).
https://doi.org/10.1016/j.jvoice.2018.03.003

- Manuscripts **III** and **IV** were based on three months research stay (5.09.2017 to 4.12.2017) of the author at the Department of Phoniatrics, University of Helsinki, Finland, supported by the funds from **Erasmus Plus** program (grant no. 2017–2018/164, Palacký University Olomouc, Czech Republic).

- Manuscripts **I** and **II** were supported by the Technology Agency of the Czech Republic (**TAČR**) project no. TA0401087 and manuscripts **II, III** and **IV** were supported by The Czech Science Foundation (Grantová Agentura České Republiky-**GAČR**) project no. GA16-01246S.

ABSTRACT

Standard voice assessment protocols incorporate evaluating voice characteristics at respiratory, phonatory, and resonatory levels. The procedures use both subjective and objective assessment measures. The present thesis is based on four publications dealing with some of these assessment procedures. Specifically, the vocal fold vibration characteristics are evaluated using videokymography (VKG, a high-speed video imaging technique) in the first part of the thesis, and in the second part, teachers' voices are evaluated using perceptual and acoustic voice measures.

In the first part of the thesis, the **clinical value of the VKG** method as an additional tool to stroboscopy (gold standard) was evaluated for diagnosing and treating various voice disorders (**manuscript I**). An exploratory questionnaire was designed for this purpose and used to evaluate results of examination of outpatients in a laryngology department. The results showed that VKG was useful in 95 % of cases, by either confirming the stroboscopic diagnosis (in 31 % cases), or making the diagnosis more accurate (44 %) or adjusting the treatment recommendations (20 %). After VKG examination, the diagnostic confidence improved in 68 % cases. VKG provided insights into the vibration characteristics of the vocal folds, and helped the clinicians to take some important diagnostic and treatment decisions when the diagnosis based on stroboscopy was uncertain. Analysis of the results also showed that the shape of lateral peaks (sharp versus rounded) and missing/reduced mucosal waves were the most helpful visual features in VKG images for obtaining the final diagnosis and providing insights into the health and pliability of vocal fold mucosa. The shape of the lateral peaks in VKG images was therefore in focus of the **manuscript II** which aimed at **quantifying the sharpness of the lateral peak** using automatic image analysis methods. Open Time Percentage Quotients (OTQ) and Plateau Quotients (PQ) were defined as two types of parameters which were expected to capture the shape of the lateral peaks in the vocal fold contours. The OTQ parameters were derived as fractions of the period and PQ as a fraction of the open phase during which the vocal folds displacement exceeded a pre-determined percentage of the vibratory amplitude. Results showed that the OTQ and PQ parameters derived at 95 % (OTQ_{95}, PQ_{95}) and at 80 % of the vocal fold amplitude (OTQ_{80}, PQ_{80}), had strong and significant correlations with the visual ratings of lateral peak sharpness (P<0.001). Therefore these quotients were considered to be the best objective parameters for quantification of lateral peak sharpness. The quotients increased their values when the shape of the lateral peak changed from sharp to round.

In the second part of the thesis, teachers' voices were investigated using **perceptual (subjective) evaluation and acoustic (objective) voice measures**. Teachers are sensitive to voice attrition as a result of teaching in vocally demanding conditions and therefore investigations of teachers are important for better understanding of the voice properties and conditions influencing teachers' voice quality. In **manuscript III**, the participants were Finnish teachers who had no voice

complaints, but in some of them laryngeal pathology was detected laryngoscopically. The changes in the values of non-smoothed and smoothed cepstral peak prominence (CPP & CPPS), voice sound pressure level (SPL) and perceptual ratings (vocal quality and vocal firmness) for three voice tasks (comfortable vowel phonation, comfortable and loud speech samples) were investigated and the influence of laryngeal pathology on these measures was studied. The results showed that CPP, CPPS and SPL values were significantly higher for vowels and loud speech than for comfortable speech ($P<0.001$). Significant correlations were found between SPL and cepstral measures. Loud speech was perceived to be firmer and have a better voice quality than comfortable speech. No significant relationships of the laryngeal pathology status with cepstral values, perceptual ratings, or voice SPL were found, however ($P>0.05$). It was concluded that neither the acoustic measures (CPP, CPPS, and SPL) nor the perceptual evaluations could clearly distinguish between healthy and disordered larynges when the pathologies are not self-perceived negatively by the teachers. Considering no vocal complaints of the subjects, the acoustic data are considered representative for teachers with functionally healthy voice.

In the final **manuscript IV**, the **influence of noise in classrooms, resulting from inappropriate location (or site) and classroom conditions, on Egyptian teachers' voices were investigated.** The results showed that there were significant correlations ($P<0.05$) between teachers' voice symptoms (severe dysphonia, neck pain, and increased vocal effort with weekly or daily recurrence) and the reported noise resulting from poor classroom conditions (overcrowded with students and poor classroom design) as well as inappropriate school and classroom locations (near road traffic). This necessitates solutions for the future improvement of conditions of schools and classroom in Egypt and elsewhere, considering vocal and general health of teachers.

ABSTRAKT

Standardní postupy vyšetření hlasu zahrnují hodnocení charakteristických vlastností hlasu na úrovních respirace, fonace i rezonance. Taková vyšetření využívají jak subjektivních tak objektivních měření. Tato dizertace je postavena na čtyřech autorských publikacích, které se věnují vybraným typům vyšetření. Konkrétněji, v první části práce jsou hodnoceny charakteristiky kmitání hlasivek pomocí vysokorychlostní zobrazovací metody videokymografie (VKG) a v druhé části práce je hodnocen hlas učitelů pomocí percepčních a akustických metod.

První část práce se věnuje **klinickému významu videokymografie**, jakožto doplňkové metody k videostroboskopii, pro diagnostiku a léčbu poruch hlasu (**publikace I**). Pro tento účel byl vyvinut speciální dotazník, který byl používán pro hodnocení výsledků vyšetření pacientů v laryngologické ambulanci. Výsledky ukázaly, že videokymografie byla užitečná v 95 % případů, kdy toto vyšetření vedlo buď k potvrzení původní stroboskopické diagnózy (31 %), k upřesnění diagnózy (44 %), nebo ke změně doporučeného postupu léčby (20 % případů). Jistota diagnózy vzrostla po tomto vyšetření u 68 % případů. Videokymografie poskytla podrobnější informace o způsobu kmitání hlasivek a pomohla učinit důležitá diagnostická a terapeutická rozhodnutí zejména v případech, kdy původní stroboskopická diagnóza byla nejasná. Analýza výsledků také ukázala, že za nejužitečnější vizuální rysy kmitání hlasivek ve videokymografii byly považovány tvary laterálních vrcholů kmitů (ostré versus zakulacené) a chybějící či redukované slizniční vlny, které jsou indikátory poddajnosti a zdraví sliznice hlasivek. **Tvar laterálních vrcholů kmitů hlasivek** byl hlavním tématem následné **publikace II**, jejímž cílem bylo nalezení objektivního parametru pro kvantifikaci ostrosti těchto vrcholů metodami obrazové analýzy. Pro tento účel byly definovány dva druhy parametrů – koeficienty relativní doby otevření (Open Time Percentage Quotients, OTQ) a rovinné koeficienty (Plateau Quotients, PQ). Tyto koeficienty byly definovány jako poměr času, po který byla výchylka hlasivky větší než stanovené procento (kritérium) amplitudy kmitů, vůči periodě (OTQ) nebo vůči době otevření hlasivek (PQ). Porovnání s vizuálním hodnocením ostrosti tvaru laterálních vrcholů ukázalo statisticky nejvýznamnější korelace (P<0.001) pro koeficienty s kritériem 95 % amplitudy kmitů (OTQ_{95}, PQ_{95}) a 80 % amplitudy kmitů (OTQ_{80}, PQ_{80}). Na základě těchto výsledků byly koeficienty OTQ_{95}, PQ_{95}, OTQ_{80} a PQ_{80} určeny jako nejvhodnější pro kvantifikaci ostrosti tvaru kmitů hlasivek z videokymografie. Tyto koeficienty zvyšují své hodnoty při změně tvaru z ostrého na zakulacený.

Druhá část dizertace se věnuje **hodnocení hlasu učitelů percepčními (subjektivními) a akustickými (objektivními) metodami.** Učitelé jsou náchylní na opotřebení hlasu, neboť pracují v hlasově náročných podmínkách a vyšetřování učitelů je proto důležité pro porozumění vlastnostem hlasu a podmínkám, které kvalitu hlasu učitelů ovlivňují. **Publikace III** studuje hlas finských učitelek, které si nestěžovaly na hlasové problémy, ale u některých z nich byla laryngoskopicky objevena patologie hrtanu. Byly hodnoceny tři hlasové projevy (pohodlné

vyslovování samohlásek, pohodlné čtení a hlasité čtení) a projev patologie hrtanu na vlastnostech hlasu. Objektivně byla hodnocena prominence špiček hlazeného a nehlazeného kepstra (non-smoothed & smoothed cepstral peak prominence, CPP & CPPS) a hladina akustického tlaku hlasu (voice sound pressure level, SPL). Percepčně byla hodnocena kvalita a pevnost hlasu. Výsledky ukázaly statisticky významně vyšší hodnoty parametrů CPP, CPPS and SPL pro samohlásky a pro hlasité čtení než pro pohodlné čtení (P<0.001). Mezi SPL and kepstrálními hodnotami byly statisticky významné korelace. Hlas byl u hlasitého čtení percepčně hodnocen jako kvalitnější a pevnější než u pohodlného čtení. Žádný akustický ani percepční parametr nevykázal statisticky významné změny vlivem patologie hrtanu (P>0.05), což naznačuje, že žádný z těchto parametrů není schopen rozlišit mezi hlasy s patologií a bez patologie hrtanu, pokud tyto patologie nejsou negativně vnímány samotnými učiteli. Obdržené výsledky je možno považovat za reprezentativní pro učitelky s funkčně zdravým hlasem.

Poslední publikace (**publikace IV)** studuje **vliv hluku v třídách v Egyptě, způsobených nevhodnou polohou a nevhodnými podmínkami, na hlas učitelů.** Výsledky prozradily statisticky významné korelace (P<0.05) mezi symptomy učitelů (závažná dysfonie, bolest krku, zvýšené hlasové úsilí s týdenním či denním opakováním) a uváděným hlukem způsobeným špatnými podmínkami ve třídách (příliš mnoho studentů, špatná konstrukce tříd), ale i nevhodnou polohou škol a tříd (hlučné oblasti). Tato situace vyžaduje nalezení řešení pro zlepšení podmínek škol a tříd v Egyptě a jinde, tak aby byla zajištěna zdravá vokální komunikace a obecně zdraví učitelů.

1 MOTIVATION FOR THE WORK OF THE AUTHOR

Videokymography (VKG) (an instrument used to directly visualize the vocal fold vibrations) was first introduced to me when it was procured for the first time in India (Kasturba Medical College, Mangalore, Karnataka, Manipal University) while I was pursuing my master's degree in Audiology and Speech Language Pathology (MASLP). Fascinated by its applications in assessing the vocal fold vibratory behavior, and its superiority to other existing visual techniques, I developed interest to learn more about it, and that's why I came to Palacký University Olomouc, Czech Republic for my doctoral studies under the guidance of Dr. Jan Švec, the inventor of VKG himself. Initially, investigating vocal fold vibrations using VKG and finding its clinical value was the primary theme of my PhD studies. However, during my course of study, I also got an opportunity to have my three months research stay at University of Helsinki in Finland through the Erasmus plus student exchange program. Under the supervision of Dr. Ahmed Geneid and Dr. Anne-Marie Laukkanen (University of Tampere, Finland), who were already working on evaluating the factors affecting teachers' voice and on perceptual and acoustic evaluation of voice of teachers, I learnt more about these topics, and became involved in the projects, which resulted in two additional papers related to teachers' voice that I authored.

The present thesis is therefore divided into two parts. The first part focuses on investigating the clinical value of VKG in assessing and treating various voice disorders, and quantifying the 'sharpness of lateral peak' parameter of vocal fold vibration using VKG images. The second part of the thesis focuses on investigating teachers' voice through perceptual evaluation and cepstral analysis (an acoustic voice measure) and assessing the influence of noise resulting from unfavorable environmental conditions (inappropriate location of school and poor classroom conditions) on teachers' voice.

- The following introduction section provides a brief description of the mechanisms of vocal fold vibration as well as of the voice disorders and their classification systems. A note on professional voice users, particularly the teacher's voice and factors affecting the teachers' voice is also provided. Furthermore, this section provides a description of the recommended protocols for assessment of voice disorders. The concepts related to auditory perceptual, visual and objective evaluation of voice is described.

2 INTRODUCTION

2.1 Voice Production Mechanism

Vocal folds are paired structures, integral for voice production. Structurally, the vocal folds are comprised of mucosa (which is normally soft, fluid-like and pliable), the non-muscular ligament and a muscle layer (**Figure 1**).

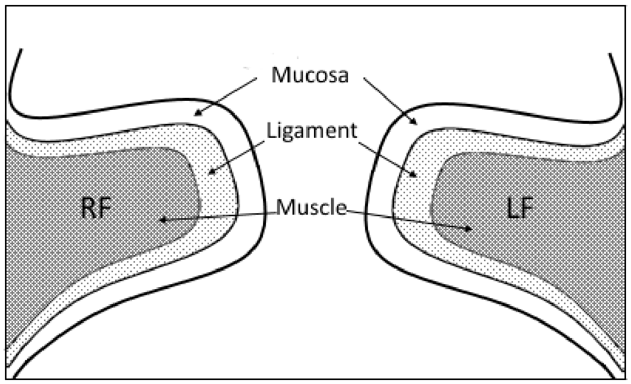


Figure 1: Coronal section through the right (RF) and left vocal fold (LF), showing the different tissue layers

When the vocal folds vibrate, they are not just opening and closing but follow a series of dynamic events along the vertical dimension of the vocal folds. The mucosal tissue of the vocal folds support the propagation of surface waves, called *'mucosal waves'* (analogous to the waves on the surface of water) occurring on the vibrating vocal folds. These waves first originate at the inferior surface of the vocal fold mucosa, propagating vertically along the medial surface and then horizontally along the superior surface of the vocal folds. This wave like motion is associated with the lower margins (or edges) of the vocal folds opening and closing earlier than the upper margins creating a delayed movement between them, a phenomenon called *'vertical phase difference'*. The vertical phase difference facilitates the delivery of airflow energy to the vocal fold tissues for sustaining oscillations of the vocal folds (McGowan, 1990, Titze et al., 1993, Titze, 1988). **Figure 2** depicts eight phases of an oscillatory cycle of the vocal folds: In the first phase, the lower margins of the vocal folds start to open. Here the vocal folds appear to remain closed when viewed from above, because the upper margins still remain closed. In the second phase, the upper margins also start to open. In the third phase, both the lower and upper margins are opening. During the fourth phase, the lower margins attain their maximal open position, while the upper margins are still opening. In the fifth phase, the lower margins are closing and are visible because upper margins attain their maximal open position. In the sixth phase, both the lower and upper margins are closing, and the mucosal waves propagate laterally on the upper surface of the vocal folds (see the outward arrows on the vocal fold surface in **Figure 2a (6)**). During the seventh phase, the lower margins completely close, followed by the complete closure of the upper margins in the final eighth phase, thus completing one cycle of vibration (Švec et al., 2009).

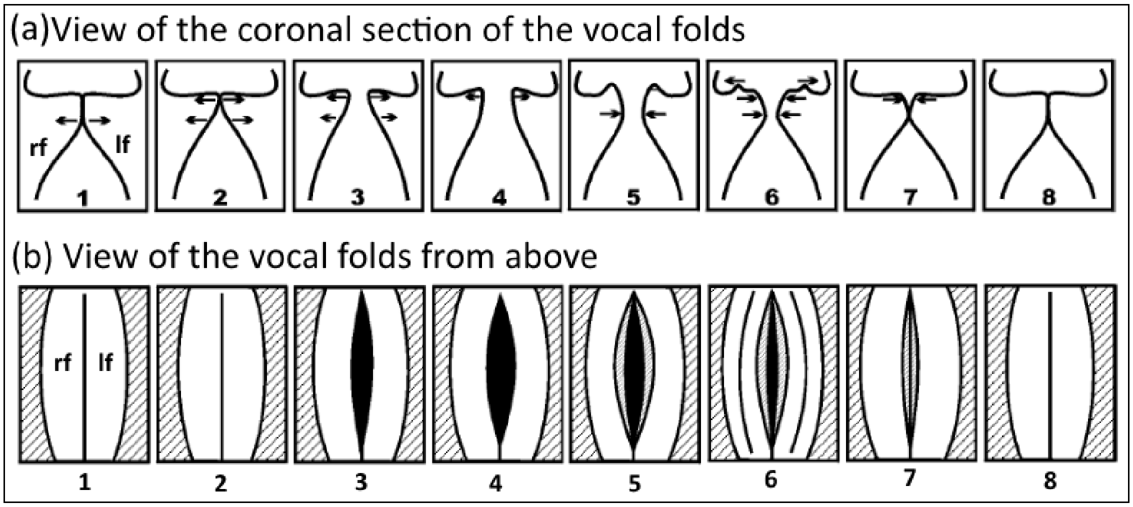


Figure 2: Schematic illustration of eight phases of an oscillatory cycle of right (rf) and left (lf) vocal folds in (a) coronal view and (b) top view of the vocal folds. Figure modified from Švec et al. (2009).

When the vocal folds vibrate, they modulate the airflow from the lungs into a series of airflow pulses to produce sound. This sound (small changes in the air pressure) propagates through the vocal tract, gets filtered acoustically, and is then radiated out through the mouth (and sometimes also through the nose) as voice (Herbst and Švec, 2016, Zhang, 2016).

The mechanism of voice production (or vowel production in particular) has been explained based on *'source-filter theory'* formulated by Fant (1960). This theory can be explained based on both (a) time and (b) frequency domain representations as shown in **Figure 3** (Story, 2002). In the time domain representation (**Figure 3a**), the sound source or the glottal flow (obtained by converting the steady airflow from the lungs into a series of airflow pulses by the vibrating vocal folds) propagates through the vocal tract filter (which is represented in **Figure 3** as a vocal tract area function), and gets transformed into the output pressure. The vocal tract area function refers to the cross-sectional area of the vocal tract as a function of distance from the glottis. In the frequency domain (**Figure 3b**), the output spectrum is the product of source spectrum and the frequency response of the vocal tract. The source spectrum here represents the spectrum of glottal flow, consisting of the fundamental frequency and its integral multiples or harmonics, and the transfer function is the vocal tract resonances or formants.

The radiated output spectrum also includes the radiation characteristics of the mouth (not shown in **Figure 3b**). When the sound escapes from the mouth, it radiates out into the free space spreading in all directions. The mouth acts as a high-pass filter, radiating more energy in higher frequencies than in lower frequencies. According to the source-filter theory, the radiated output spectrum of voice is the product of the source spectrum, the vocal tract transfer function, and mouth radiation, mathematically represented in equation (1) described by Kent and Read (1992).

$$\mathbf{P(f) = U(f) \times T(f) \times R(f) \dots\dots\dots (1)}$$

where P(f) is the radiated sound pressure spectrum of voice, where (f) indicates function of frequency, U refers to the air volume velocity, as the vocal folds act as a source of air pulses, T is the vocal tract transfer function, and R is the mouth radiation. The output voice signal, in form of acoustic pressure, can be captured by a microphone. The microphone converts the acoustic signal into electrical signal, which can be used for further analysis of fundamental frequency, vocal intensity and other voice parameters.

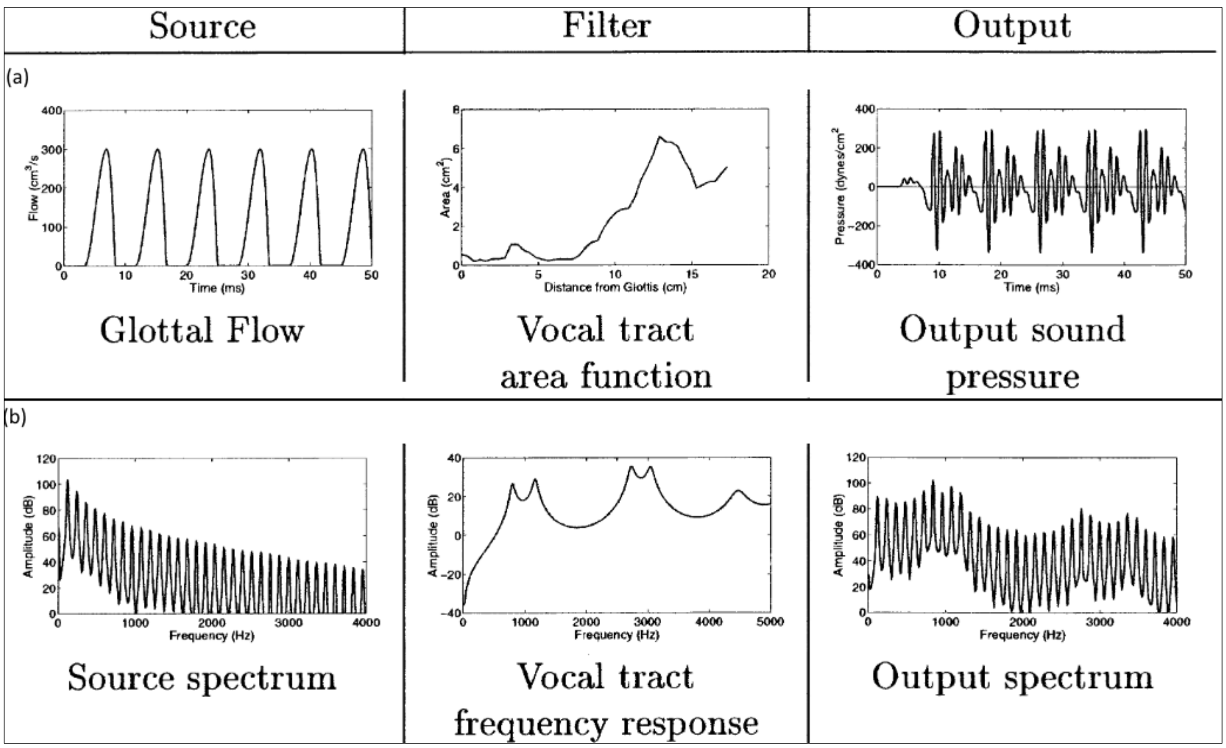


Figure 3: Illustration of source-filter theory in (a) time domain and (b) frequency domain representation. The first row shows the source waveform, vocal tract area function, and the output pressure waveform, all in time domain and the second row shows their corresponding quantities in frequency domain. Figure modified from Story (2002).

2.2 Voice disorders and their classification

Individuals with voice disorders may present with either absence of voice (aphonia) or have a variable degree of vocal impairment (dysphonia) at the level of respiratory, laryngeal (phonatory) and/or resonatory systems. Dysphonia is synonymously used with the term *'hoarseness'* and is defined as a disorder with alterations in the voice quality, pitch (perceptual correlate of fundamental frequency of voice (fo)), loudness (perceptual correlate of vocal intensity) or vocal effort that disrupts communication and affects the voice related quality of life of an individual (Schwartz et al., 2009). A wide array of voice disorders occurs in all age groups (from pediatrics to geriatrics), that may be acquired or congenital (since birth).

A comprehensive *'Diagnostic Classification System For Voice Disorders'* has been proposed with high levels of inter-rater reliability (Baker et al., 2007). In this classification system, the voice disorders are broadly classified as organic voice disorders (OVD) and functional voice disorders (FVD). Individuals with OVD present with aphonia or dysphonia due to some laryngeal lesions, that changes the normal structure of the vocal folds (or associated structures), or there may be interruption to the neurological innervation of the laryngeal mechanism. FVD on the other hand refers to individuals having aphonia or dysphonia without any underlying organic lesion, or if present it may be insufficient to manifest a voice disorder, or may be considered secondary to the functional problem. The FVD is further classified into psychogenic voice disorder and muscle tension voice disorder (MTVD). In MTVD, there is misuse or dysfunction of the laryngeal musculature due to disturbed psychological process. Over the time, the abnormal vocal behaviors may manifest secondary organic lesions such as vocal nodules, resulting in a more misuse or abusive behavior of voice.

Numerous other voice disorder classification systems exist, for example the one provided by Verdolini et al. (2006) called the *'Classification Manual For Voice Disorders-I'*. This classification system uses a standard framework for classifying each voice disorder based on the following criteria: essential and associated features; vocal impairment; clinical history and demographic profile; course of the disorder and complications; medical and vocal differential diagnosis; and criteria of disorder severity. These criteria help clinicians to make an appropriate diagnosis and differentially diagnose various voice disorders. Titze (2000) has also classified voice disorders based on the responses of a biomechanical oscillator to environmental, systematic, or traumatic conditions. This classification system has four divisions: Congenital (structural) voice disorders, disorders related to tissue changes, disorders related to neurological or muscular change, and vocal fatigue.

Recently, the World Health Organization (WHO, 2016) published the ICD-10 version 16, which is the 10^{th} revision of the International Statistical Classification of Diseases and Related Health Problems. It provides diagnostic classification standard for all clinical and research purposes, with codes for diseases, causes, signs and symptoms of injury or diseases. **Table 1** provides information on the classification of disorders leading to voice problems, used in **manuscript I** of the present thesis.

Table 1: ICD-10 Version: 2016 classification of voice disorders and related diseases

ICD-10 Code	Disease Category	Sub-code	Structures involved and disease subtypes
C32	Malignant neoplasm of the larynx (neoplasm- abnormal growth of tissue)	C32.0	Glottis (true vocal cords*)
		C32.1	Supraglottis: Aryepiglottic fold, Epiglottis (suprahyoid portion), False vocal cord, Posterior (laryngeal) surface of epiglottis
		C32.3	Subglottis
D02	Carcinoma in situ	D02.0	Larynx Epiglottis (suprahyoid portion)
D14	Benign neoplasm of the larynx	D14.1	Also papillomatosis of the larynx
J04	Acute laryngitis and tracheitis	J04.0	Acute laryngitis: Oedematous, Subglottic, Suppurative, Ulcerative
		J04.1	Acute tracheitis , Catarrhal (inflammation of mucous membrane)
		J04.2	Acute laryngotracheitis
J37	Chronic Laryngitis	J37.0	Catarrhal (inflammation of mucus membrane, Hypertrophic, Sicca (in Latin, *siccus*, meaning 'dry')
		J37.1	Chronic laryngotrachetis
J38	Diseases of vocal cords and larynx, not elsewhere classified	J38.0	Paralysis of vocal cords and larynx Laryngoplegia
		J38.1	Polyp of vocal cord
		J38.2	Nodules of vocal cords: Chorditis (fibrinous) (nodosa) (tuberosa) Singer nodes and Teacher nodes
		J38.3	Other diseases of vocal cords: Abscess, Cellulitis, Granuloma, Leukokeratosis, Leukoplakia
		J38.4	Oedema of larynx (Oedema (of): glottis, sub and supraglottis)
		J38.5	Laryngeal spasm, Laryngismus (stridulus)
		J38.6	Stenosis of larynx
		J38.7	Other diseases of larynx: Abscess, Cellulitis, Necrosis, Pachyderma, Perichondritis, Ulcer
J39	Other diseases of upper respiratory tract		
J40	Bronchitis, not specified as acute or chronic		
J42	Chronic bronchitis Chronic tracheitis or tracheobronchitis		
K21	Gastro-oesophageal reflux disease and Laryngopharyngeal reflux disorder (LPRD)		

*The terms 'vocal cords' and 'vocal folds' are used synonymously. The vocal folds behave like the strings/cords of guitar or violin and therefore the term 'vocal cord' is sometimes used.

2.3 Professional voice users: Teachers' voice

Any individual, whose professional (or employment) activities are dependent on efficient use of voice, is considered as a professional/occupational voice user. Four levels of voice use (Koufman and Isaacson, 1991) have been categorized based on the significance of voice to work-related or professional activities:

i. Level I- Elite Vocal Performer–Are *singers and actors* to whom, even a slight aberration of voice may lead to dire consequences in their profession.
ii. Level II- Professional Voice User–Are *clergy, teachers, lecturers, receptionists* to whom a moderate voice problem might prevent adequate job performance.
iii. Level III- Non-Vocal Professional–Are *lawyers, physicians, businessmen, etc.* to whom only a severe voice deviation may affect adequate job performance.
iv. Level IV- Non-Vocal Non-Professional–Include *clerks, laborers, etc.*to whom vocal quality is not a prerequisite for adequate job performance.

Professional voice users demand more attention than non-professional voice users, as they often tend to overuse their voices without appropriate voice training and vocal hygiene. One of the largest groups of professional voice users are teachers. Teacher's voice is vulnerable to disorders as a result of prolonged voice use in heavy vocally loading conditions (Vilkman, 2000). In a systematic review, Cutiva et al. (2013) reported results from 23 publications focusing on the prevalence of voice disorders in teachers from the American and European countries. Their review showed that the life-time prevalence of voice disorders in teachers ranged between 51 % (Angelillo et al., 2009) to 69 % (Sliwinska-Kowalska et al., 2006) and the prevalence of voice disorders with unspecified recall period ranged between 13 % (Jónsdottir et al., 2002) to 94 % of teachers (Roy et al., 2004).

2.3.1 Factors affecting teachers' voice

In general, the major risk factors associated with occupational voice disorders can be divided into two broad categories: ergonomic (or environmental) and extra-occupational (or individual) risk factors as shown in **Table 2** (Morawska and Niebudek-Bogusz, 2017).

Factors such as poor environmental conditions (Vilkman, 2000, Rantala et al., 2015, Cutiva et al., 2017, Durup et al., 2017), poor working conditions (Kankare et al., 2012, Cutiva et al., 2013), unawareness of appropriate vocal hygiene (Bolbol et al., 2017) and lack of voice training (Ilomäki et al., 2005) have been studied and attributed to development of voice disorders in teachers.

Table 2: Two broad categories of risk factors associated with occupational voice disorders

Ergonomic (environmental) risk factors	**Extra-occupational (individual) risk factors**
Vocal loading	Incorrect voice technique
Work related stress	Extra-occupational voice activities
Poor working posture	Co-existing disorders (inflammatory diseases of respiratory airways, hormonal disorders, reflux etc.)
Poor Air quality, dryness, dust	Personality/anxiety disorders
Background noise	Lifestyle habits (smoking, caffeine, alcohol intake)
Poor room acoustics	

Teachers are required to have a good vocal quality for effective conversations with students, as it has been reported that there could be negative influences on children's ability to process speech due to voice problems (mild or severe) in teachers (Rogerson and Dodd, 2005). According to Nelson (1999), in his study teachers and students conversed at least 60 % of the time in an active classroom scenario, stressing on the need for a favorable listening environment, allowing clear communication. If teachers present with voice problems, it could further lead to, disinterest and dissatisfaction in job, lack of self-esteem, and fatigue (Vilkman, 2000). "Bad classroom acoustics" has been indicated to be one of the greatest threats to vocal health in teachers (Vilkman, 2000). Teachers raise their voices in presence of noise that leads to increased vocal fatigue and stress in teachers (Tiesler and Oberdörster, 2008). Classroom noise, reverberation, echoes, all interfere with the ability of the students to understand speech, resulting in an increase of vocal effort by teachers (Berg et al., 1996).

2.3.2 Voice symptoms in teachers

Extensive research has been carried out in studying the voice problems and vocal symptoms in teachers with various teaching levels (kindergarten to primary and secondary schools), based on work and employment conditions (topics of teaching- music, academics, physical trainers), and individual factors such as voice use, number of years of teaching, and other psychosocial aspects (Kankare et al., 2012, Thibeault et al., 2004, Munier and Kinsella, 2007, Sliwinska-Kowalska et al., 2006, Abo-Hasseba et al., 2017, Morrow and Connor, 2011, Cutiva et al., 2013). Most of these are questionnaire studies that evaluate frequency and severity of voice symptoms which are self-reported by the teachers. The prominent voice symptom is vocal fatigue occurring due to vocal overload (Kankare et al., 2012, Ilomäki et al., 2005, Simberg et al., 2005, Sala et al., 2001). Vocal fatigue symptom occurring weekly has led to vocal activity limitation and participation restriction in teachers (Ilomäki et al., 2017). **Table 3** lists various voice symptoms commonly seen in teachers.

Table 3: List of voice symptoms commonly reported by teachers

1	Hoarseness and breathiness of voice	9	Pain in the neck region
2	Voice tiredness/vocal fatigue	10	Sensation of lump in throat
3	Difficulty projecting the voice	11	Excess mucus secretion
4	Vocal discomfort	12	Frequent throat clearing
5	Increased effort to talk	13	Strained voice quality
6	Chronic throat dryness or soreness	14	Voice instability
7	Trouble speaking or singing	15	Aphonia
8	Presence of voice breaks		

2.4 Recommended protocols for assessment of voice disorders

The foremost step in designing an efficient management program for voice disordered individuals is evaluating the voice disorder characteristics and the effect that disorder has on a person's potential to communicate. When assessing the voice characteristics, the clinicians follow a standard voice evaluation protocol, where the voice is evaluated carefully and systematically at various levels, i.e., at respiratory, phonatory, and resonatory level, which all participate in a coordinated fashion for the production of voice.

In 2001, the European Laryngological Society (ELS) proposed a basic voice assessment protocol for improving the methodology of the functional assessment of vocal pathology. Its main aim was to enable optimum comparisons with literature and enable publishing consistent and reproducible results of any kind of voice treatment. This assessment protocol aims at having comparable results of voice assessment, particularly before and after performing phonosurgery (Dejonckere et al., 2001). It describes five sets of voice measurements for assessing most 'common dysphonias' including the (1) auditory-perceptual evaluation, (2) laryngeal examination (videostroboscopy), (3) acoustic evaluation, (4) aerodynamic/efficiency evaluation, and (5) subjective (self-) evaluation of voice by patient. More recently, the American Speech and Hearing Association (ASHA) expert panel further elaborated instrumental parts of the ELS assessment protocol and developed a standard protocol for instrumental analysis of vocal function (Patel et al., 2018). It provides detailed description about the technical and procedural information about the instrumental assessments including (1) laryngeal endoscopic imaging techniques, (2) acoustic and (3) aerodynamic procedures. These recommendations are put forth in order to have an (a) evidence-based practice guidelines for valid and reliable voice assessment procedures, (b) to have comparable assessment results across various research and clinical facilities and (c) facilitate treatment efficacy.

- The following sections will briefly describe: (1) auditory-perceptual evaluation, (2) self-evaluation of voice by patient, (3) laryngeal examination, and (4) acoustic evaluation of voice. Information on the previous and current work done related to only these assessment procedures are provided here, as these are most relevant to the published papers in the present thesis. The ELS and ASHA recommendations are also described briefly under each of these procedures.

2.5 Auditory-perceptual evaluation of voice

Auditory-perceptual evaluation of voice is considered to be the most common and gold-standard procedure for documenting voice characteristics of an individual (Oates, 2009). Since the 'voice quality' is merely perceptual in nature, the success and effectiveness of voice management is based on how normal the voice sounds. To make the perceptual evaluation quantitative and more meaningful, listeners judge the voice based on some standard rating scales such as Grade, Roughness, Breathiness, Aesthenia, Strain (GRBAS) (Hirano, 1981b), Roughness, Breathiness and Hoarseness (RHB) (Nawka et al., 1994), Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) (Kempster et al., 2009) etc. Various other voice quality types are described in literature, such as a pressed voice (tight vocal fold adduction leading to increased vocal fold impact stress) (Jiang and Titze, 1994), resonant voice (a voice produced with ease, vibrant in facial tissues, whose quality is neither pressed nor breathy) (Titze, 2001), and firm voice, etc. These voice quality types and the general voice quality are also rated perceptually on different types of rating scales. For example, Järvinen et al. (2017) in their study rate the general voice quality along a ten point unipolar scale from 0 = very poor to very good = 10 and firmness of voice along a scale from 0 = very pressed to very breathy = 10. There is a need for standardization of such rating scales.

ELS (Dejonckere et al., 2001) in their basic voice assessment protocol recommend the use of GRB scale (Grade, Roughness & Breathiness) or RBH scale rated either by using a four-point rating scale (0-normal or absence of voice deviance; 1-slight deviance; 2-moderate deviance; 3-severe deviance) or using a visual analogue scale. A visual analogue scale consists of an undifferentiated line, often 10 cm long, on which listeners assign a value to the voice sample they hear thus indicating the quality of voice.

Although perceptual evaluation has been the gold standard assessment procedure, it has been criticized for its reliability, as it is subjective and depends highly on listeners' internal standards, listeners' experience, type of rating scales used and type of voice samples used such as sustained vowel, conversational speech, standard reading text, picture description, etc. (Oates, 2009). Therefore, for a reliable documentation of voice characteristics, instrumental analysis of voice has been used. Nevertheless, listeners' judgments remain a standard against which other objective voice measures (acoustic, aerodynamic, etc.) are usually compared and evaluated.

- Auditory-perceptual evaluation of voice quality and vocal firmness rating has been used in the **manuscript III** of the present thesis.

2.6 Self-evaluation of voice by patient

Auditory-perceptual evaluation of voice is usually physician-driven and forms an essential part of voice assessment protocol. However, self-evaluation of voice by patient may also be equally important and relevant for diagnostic purposes as well as act as a valid measure to test the treatment effectiveness. Patients judge the success of treatment based on the improvements in their voice and their perception of voice abnormality.

ELS (Dejonckere et al., 2001) in their basic voice assessment protocol for self-evaluation of voice recommend the use of Voice Handicap Index (VHI) to document the quality of life of a voice disordered individual, based on functional, emotional and physical impact of voice (Jacobson et al., 1997, Rosen et al., 2004, Zur et al., 2007, Niebudek-Bogusz et al., 2011, Arffa et al., 2012). The VHI-30 (which includes 30 items) and VHI-10 (shortened version with 10 items) questionnaires are rated on four point rating scale (0-never; 1-almost never; 2-sometimes; 3-almost always and 4-always), where the total score helps categorize quality of life of an individual with voice problems based on mild, moderate and severe categories. The physical component in VHI assesses various voice symptoms such as, worsening of voice at the end of the day, increased vocal effort, increased strain to speak, running out of air while talking, voice instability etc. Reporting such voice and throat symptoms by patients is an essential part of voice evaluation procedure. It may be useful to clinicians for assessing the progress of the disorder and influence of the intervention program for improvements in the voice of an individual.

- In the present thesis, for the purpose of assessing the frequency and severity of voice symptoms of teachers, self-reporting questionnaire was designed and utilized in the **IVth study**. The voice symptoms included in this questionnaire were: dysphonia, laryngeal pain, throat clearing, throat dryness, voice interrupted by the end of the day, and extra voice effort required to continue speaking. The frequency of these voice symptoms were rated by the teachers on a four-point rating scale (1 = no recurrence; 2 = monthly recurrence; 3 = weekly recurrence; 4 = daily recurrence) and the severity of voice symptoms were also rated on a four-point rating scale (1 = none; 2 = mild; 3 = moderate; 4 = severe).

2.7 Laryngeal examination (direct visualization of the larynx or vocal folds)

Direct visualization of the larynx is a prerequisite to accurate diagnosis and treatment of various voice disorders. Laryngoscopy, which is the term used to describe the visualization/examination of the larynx, has evolved from use of direct and indirect visualization techniques to advanced endoscopic and imaging techniques available today. The direct laryngoscopy involves visualizing the larynx (including the vocal folds) with a direct and unimpaired line of sight from the eye of the clinician to the patient's glottal opening. The indirect techniques on the other hand do not involve direct line of sight, but a light is either reflected off or passed through an instrument before reaching the eye of the clinician or the light sensor (Collins, 2014, Best and Akast, 2016). These indirect techniques are commonly used in awake patients and are described further.

2.7.1 Indirect laryngoscopy

2.7.1.1 Mirror laryngoscopy

Manuel Garcia Garcia (1855) was the first to visualize the human vocal folds practically by using a small bent mirror (similar to a dental mirror) placed posteriorly in the throat and angled down in the laryngopharynx. In this procedure, the subject was seated facing the sun, and the luminous rays of the sun falling on the mirror reflected on the larynx enabling one to visualize the vocal folds and associated structures. Later on Czermak (1858), as mentioned by Jahn and Blitzer (1996), Cooper (2004), Rosen et al. (2009), used an external light source instead of depending on the sun as a source of illumination. The light from the external source was incident on the perforated head mirror worn by the examiner or clinician, which was then reflected in a direction coaxial to the gaze while maintaining a binocular vision (**Figure 4**). The person undergoing this procedure was seated facing the examiner and instructed to lean forward from the waist, with chin up, and tongue sticking out as if in a 'panting' position. The subject was then asked to sustain a phonation of high-pitched vowel [i:] sound, which brought the tongue base forward, enabling visualization of vocal folds. The total time taken for this examination was only 5 to 10 minutes.

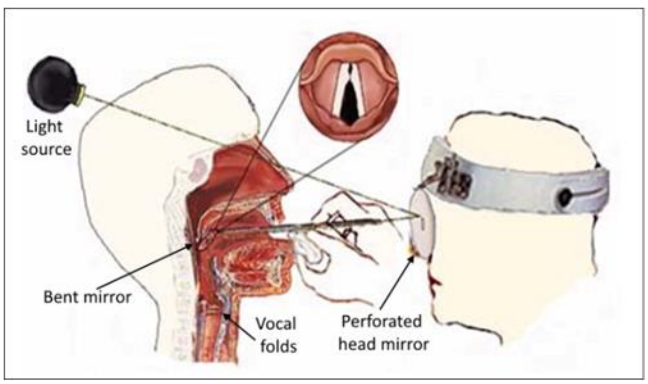


Figure 4: Image illustrating mirror laryngoscopy procedure. Figure taken from Saldanha et al. (2013).

This procedure is continued to be used in clinics even today. However, there are some advantages and disadvantages of its use. The main advantage of this method is that it is simple, inexpensive and allows a quick overview of the parts of the larynx. However, the principle disadvantage is that it lacks magnification for evaluating fine laryngeal details, such as epithelial lesions, and phonotruamatic changes. Second disadvantage is that there is no means of recording or reviewing the images during follow-up sessions for comparisons. Other disadvantages are that it is poorly tolerated due to gag-reflex, can be used only to assess sustained vowel sounds preferably [i:] and alters normal laryngeal biomechanics (Sataloff et al., 2015).

In spite of all these disadvantages, mirror laryngoscopy occupies a special place in the history of laryngoscopy. Clinicians make use of this method even in the current era of high-definition imaging. It is highly recommended to be performed during the first visit of the patient, as part of the routine clinical evaluation in order to avoid any professional diagnostic delay of voice disorders, especially in cases of laryngeal cancer (Teppo and Alho, 2008).

- In the present thesis, mirror laryngoscopy examination has been used for visualizing the laryngeal structures of teachers, in **manuscript III**.

2.7.1.2 Flexible and rigid laryngoscopy

The mirror laryngoscopy requires some manual dexterity and finesse, and so over the years this method has been supplanted by other indirect laryngeal visualization techniques such as flexible and rigid laryngoscopy. **Figure 5** shows the flexible nasoendoscope and rigid endoscope used in these techniques respectively. **Table 4** specifies the differences between the flexible and rigid laryngoscopy in terms of its equipment and technique. Also important advantages and disadvantages are listed.

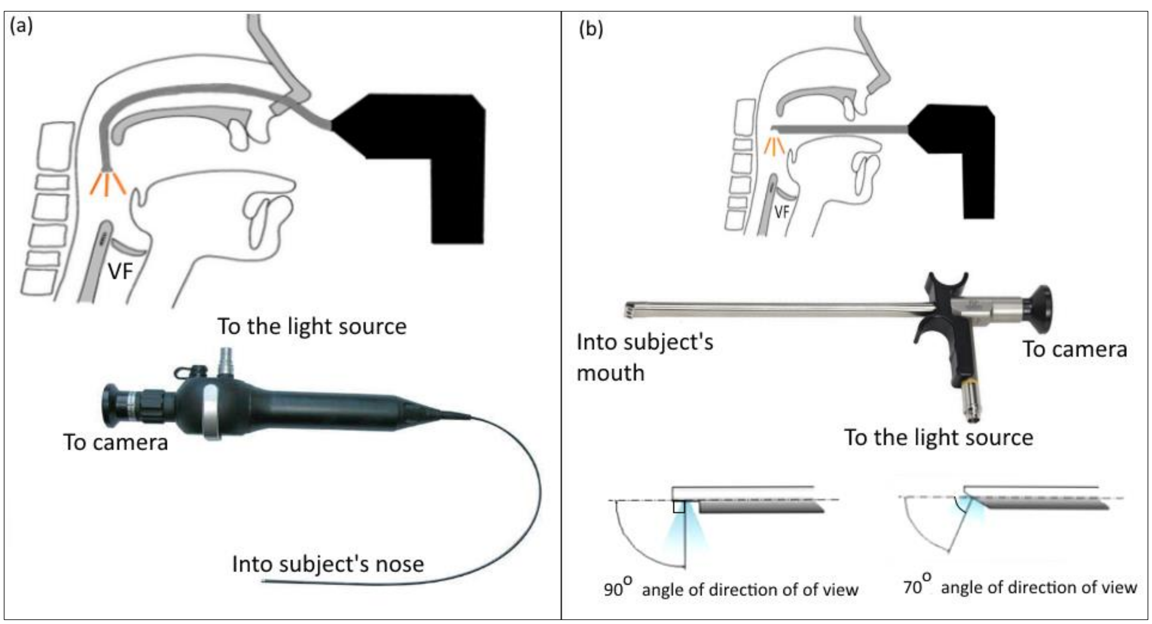


Figure 5: Visualization of vocal folds (VF) via (a) flexible laryngoscopy with flexible nasoendoscope (or laryngoscope), and (b) rigid laryngoscopy with 90° and 70° rigid endoscopes (or laryngoscopes). Figure modified from Andrade Miranda (2017).

Table 4: Differences between flexible and rigid laryngoscopy (Sataloff et al., 2015)

	Flexible laryngoscopy	**Rigid laryngoscopy**
Equipment	The classic flexible fiberoptic laryngoscope (Silberman et al., 1976) consists of an objective lens at its distal end, and a flexible insertion cord comprising of bundles of optical light conducting fibers, illuminated by a connected light source. These fibers transmit the images to a proximal objective with focusing lenses. The images are either viewed by eye, or the proximal objective is connected to a camera that allows capturing images from the laryngoscope. This is then connected to an image processor and monitor enabling multiple observers to see still images or videos. Today, the most modern flexible laryngoscopes (**Figure 5a**) use distal chip technology, where a charge coupled device (CCD) is attached to the distal end of the laryngoscope. The sensor on the distal chip placed on the laryngoscope tip captures the images electronically from its point of origin allowing for a high definition resolution of the image. The laryngoscope is connected to a light source for distal illumination, and to a camera on the proximal objective which is connected to an image processor and monitor for viewing the laryngeal structures.	A 70^{o} or 90^{o} (angles of direction of view) rigid endoscope is used, that is connected to a light source for illumination (**Figure 5b**). The equipment also consists of a camera mounted on the proximal lens of the laryngoscope that allows for image recording, viewed on a monitor.
Technique	Subject is in a seated position, facing the examiner. The flexible laryngoscope is inserted via the nasal cavity, pushed down into the pharynx and positioned just above the larynx for visualizing the vocal folds and associated structures. Topical anesthesia and decongestant in nasal cavity improves patient comfort. Nasal cavity, nasopharynx, oropharynx, hypopharynx, larynx, vocal folds, all can be visualized easily. Velopharyngeal incompetence can also be observed by making patient repeat sounds /pa-ta-ka-pa-ta-ka/ or /ka-ka-ka/. Vocal fold abduction, adduction, speech and singing tasks can be performed while laryngeal functions are evaluated.	Subject is in a seated position. Topical anesthesia is sprayed in case of gag reflex. Patient is instructed to lean forward, with chin up and tongue sticking out as in panting position, while the rigid laryngoscope is inserted over the tongue in the oropharynx. The subject is asked to sustain a phonation of vowel [i:], which moves the tongue base down to bring the larynx in view.
Advantages	1. It is well tolerated. 2. The device is portable with minimum requirements of light source. 3. It allows for running speech tasks. 4. The larynx is viewed in its natural position. 5. A closer view of epithelial changes can be observed with the distal chip laryngoscope that provides high definition images.	1. Has improved optical quality 2. Helps identify subtle vocal fold mass lesions or medial surface abnormalities. 3. There is no barrel distortion effect or vignetting. Offers sharper images with a better quality than those obtained from flexible distal chip laryngoscopy.

Disadvantages	1. When the rounded (wide angle) lens in the distal chip laryngoscope advances closer to the tissue, there is significant barrel distortion effect (causing images to spherize, that is the edges of the images look bowed or curved to human eye). There is also problem of vignetting (illumination fall-off, that is the darkening of image corners) 2. The images produced by the distal chip are processed by the image processor with its pre-determined calibration settings for color, hue and saturation. Therefore, the images produced do not necessarily depict the natural color of the laryngeal structures. This can be a disadvantage when diagnosing color dependent diagnosis of disorders such as laryngopharyngeal reflux disorder (LPRD).	1. Performing requires practice, experience 2. May be time consuming if patient is not comfortable 3. Restricted to only vowel [i:] phonation, no running speech tasks 4. May not be best suitable for observation of supraglottal hyperfunction due to alterations in tongue base movements

2.7.2 Laryngeal imaging techniques for observing vocal fold vibrations

Laryngeal imaging techniques such as videostroboscopy and high-speed imaging (high-speed videoendoscopy and videokymography) procedures help evaluating both structure and function of vocal folds, and enable having a permanent record for further follow-up sessions (Woo, 2016, Bless et al., 2009). ELS (Dejonckere et al., 2001) and ASHA (Patel et al., 2018) have recommend the use of videostroboscopy for assessment of vibratory functions of the vocal folds to determine the nature and cause of dysphonia. However, when videostroboscopy renders difficulty in assessing irregular vibrations of the vocal folds, advanced imaging techniques such as laryngeal high-speed videoendoscopy (Deliyski, 2010, Patel et al., 2008) and videokymography (Švec et al., 2009, Švec and Schutte, 1996) are recommended to be used.

- The next sections will provide details about the working principle and clinical value of videostroboscopy, high-speed videoendoscopy, and videokymography.

2.7.2.1 Videostroboscopy

The human vocal folds can vibrate between ca. 65 Hz to 1400 Hz depending on the voice type and gender of the individual (Cutler and Cleveland, 2002). Observing such high frequency vibrations through naked eye, under continuous light is not possible. In such conditions, stroboscopy technique can be employed to visualize the vocal fold vibrations in slow motion. Oertel is reported to be the first scientist to use stroboscopy for examining the vocal folds in vivo (Oertel, 1878, Oertel, 1895, Bless et al., 2009).

The stroboscopy technique involves estimating the fundamental frequency of voice in real-time, and then synchronizing it with flashes of light at frequencies 1 or 2 Hz above the fundamental frequency of vocal fold vibration. This renders illusionary slow motion vibrations of the vocal folds at 1 or 2 Hz (which is the frequency difference between the frequency of strobe flashes and fundamental frequency of voice) (**Figure 6**). More information about the stroboscopic principle can be found in Mehta et al. (2010a).

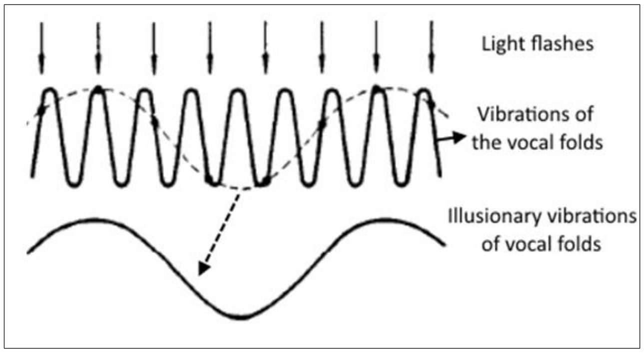


Figure 6: Illustration of the principle of videostroboscopy showing the synchronization of light flashes at frequencies slightly above the fundamental frequency vocal folds of vibration, to obtain the new illusionary slow motion of the vocal folds. Figure modified from Hirano and Bless (1993).

When the stroboscopy technique is coupled with the video camera technology for observing the vocal fold vibrations it is called videostroboscopy. This technique has been considered as a gold standard instrument for assessing various voice disorders (Hartnick and Zeitels, 2005, Patel et al., 2008) that helps clinicians to evaluate vocal fold vibratory function in real-time. However, one major limitation of this technique is that it is confined to only periodic vibrations of vocal folds. Any aperiodicity in the vibrations yields inaccurate vocal fold frequency capture making it difficult for synchronization of strobe light flashes. This may hamper the diagnostic possibilities in patients with hoarse voice quality (Patel et al., 2008). Also due to limitation in the sampling rate of the video cameras (25 or 30 images per second), it is difficult to document the vibratory characteristics of each glottal cycle. To overcome these limitations and allow for capture of multiple images per glottal cycle, the high-speed imaging techniques such as high-speed videoendoscopy and videokymography were developed. These techniques record vocal fold vibrations at higher frame rates well above the fundamental frequency of voice (Bless et al., 2009, Mehta and Hillman, 2012, Rosen et al., 2009).

2.7.2.2 High-speed videoendoscopy (HSV)

Recently, laryngeal high-speed videoendoscopy (HSV) is being increasingly used as it allows for the assessment of vocal fold vibrations especially when stroboscopy fails in cases of aperiodic vocal fold vibrations due to its advantage of high sampling rate (as high as 8000 frames per second) (Rosen et al., 2009, Deliyski and Hillman, 2010). **Figure 7** shows the comparison of imaged glottal cycles from the HSV and stroboscopy. In this figure, approximately 13 frames/images represent one glottal cycle from HSV recording, allowing for visualization and documentation of the vibratory characteristics within each glottal cycle efficiently. HSV has been reported to be better than stroboscopy in diagnosing some voice disorders (Braunschweig et al., 2008, Mehta et al., 2010b, Chen et al., 2016). However some of the disadvantages of HSV include the requirement of more advanced objective analysis procedures, for better understanding of the vocal fold movement patterns (Olthoff et al., 2007), making it more time consuming. The other disadvantages of HSV include the lack of immediate visual feedback (requires offline analysis of parameters), expensive equipment, often inferior image quality compared to videostroboscopy (due to lower spatial resolution in relation to increased recording speed) and large amount of data requiring higher storage space, thus making HSV computationally more demanding (Švec et al., 2009, Bless et al., 2009, Švec and Šram, 2011). To overcome these limitations, videokymography was developed, which will be described in more detail in next section as it is of particular interest for the present thesis in **manuscripts I** and **II**.

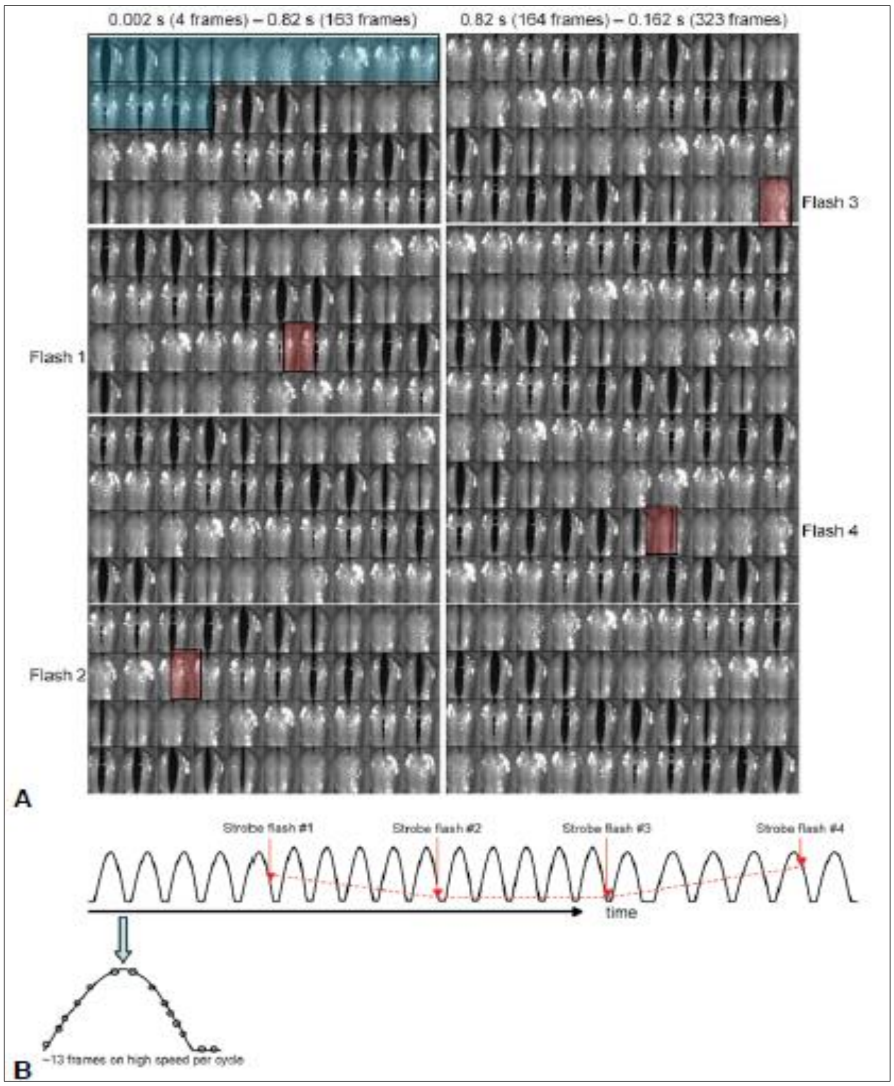


Figure 7: Comparison of imaged glottal cycles from HSV and stroboscopy. (A) Blue colored images represent one glottal cycle with approximately 13 frames from the HSV recording. The red boxes represent the images which were illuminated by the strobe light. Grey images along with the blue images are the frames not captured or missed on stroboscopy. (B) Is the representation of A, comparing glottal cycles and frames captured by stroboscopy and HSV. Figure taken from Patel et al. (2008).

2.7.2.3 Videokymography (VKG)

2.7.2.3.1 Working principle of VKG and components of VKG system

VKG (Švec and Schutte, 1996) has been developed as a simple and more economical alternative to HSV imaging technique. It is based on kymographic imaging procedure that represents the vocal fold vibrations in a single image. Dating back to 1970s, Gall and his colleagues were first to implement kymographic imaging to document vocal fold vibrations, and called it 'photokymography' that used a special photographic camera (Gall et al., 1971) to capture the vocal fold vibrations on a photographic film.

VKG is based on a special video camera that is connected to a laryngoscope. This videokymographic camera is able to record the vocal fold vibrations at a standard video rate (25 or 30 images per second) as well as at a high-speed rate of 7200 images per second simultaneously (Švec et al., 2009, Švec and Šram, 2011). The high speed imaging involves repeated scanning of the vocal fold vibrations at one horizontal line perpendicular to the glottis. The successive line images are concatenated to form a videokymogram (or kymogram), displaying the time on vertical axis, going from top to bottom as shown in **Figure 8**.

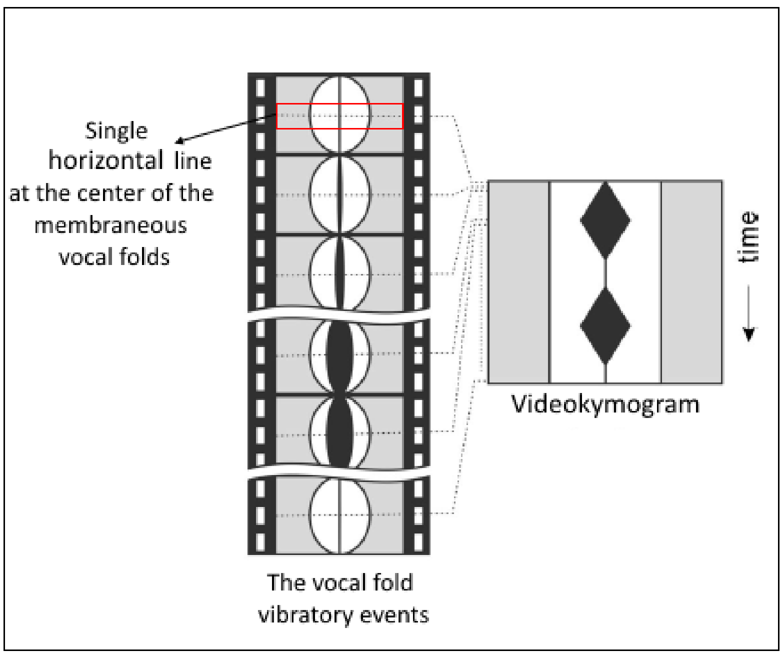


Figure 8: The videokymogram shows the vocal fold vibratory behavior, obtained by concatenating the successive line images derived by repeated scanning of one horizontal line (dotted line in red box) at the middle of the vocal folds, perpendicular to the glottis. Figure modified from Qiu and Schutte (2007).

The VKG system comprises of the following components (**Figure 9a)**: (1) a laryngoscope (70 ^{o} or 90^{o} rigid endoscope), (2) standard C-mount objective adapter for the camera, (3) a VKG camera head (consisting of charge coupled device-CCD), (4) VKG camera unit, (5) continuous light source of high intensity (preferably 300W-xenon light source, for good image quality), (6) light cable, (7) digital video-recorder or computer (VKG can be used in both the Phase Alternating Line (PAL) as well as the National Television System(s) Committee (NTSC) standards), and (8) video monitor. Microphone is needed to record also the audio signals during VKG examination.

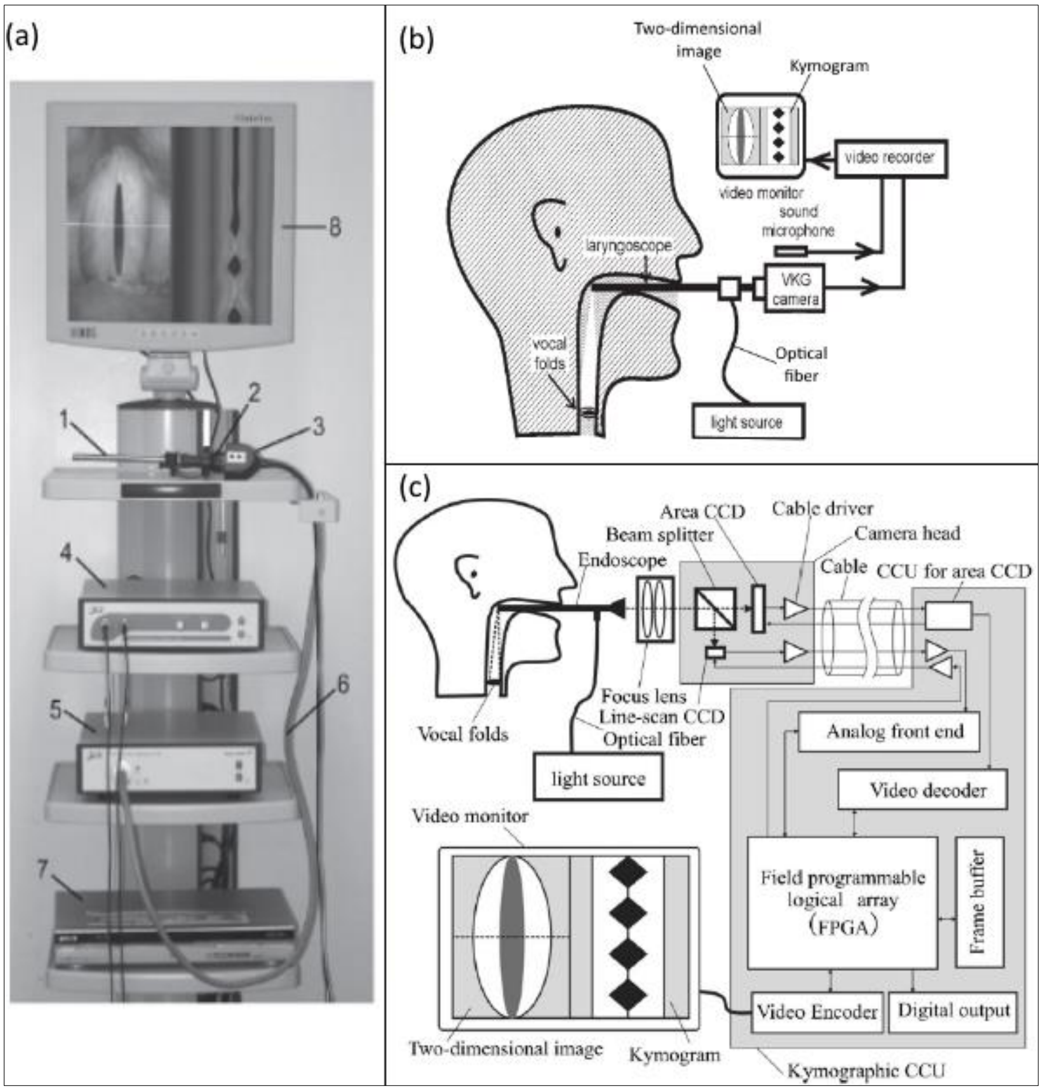


Figure 9: The components of the VKG system are shown in (a), the VKG examination set up for the patient is shown in (b), and the detailed parts of the VKG camera unit is shown in (c). See the text for description. Figure modified from Švec and Šram (2011), Qiu and Schutte (2007).

A rigid laryngeal endoscope (or laryngoscope) is inserted into the subject's mouth. It is used to visualize the vocal folds directly by the human eye (**Figure 9b)**. The laryngoscope is connected to the VKG camera head via a C-mount optical adapter. A high-intensity light source is used for illumination, which is also connected to the laryngoscope via optical fibers that transmit visible light from light source on to the vocal folds. A 300W-xenon light source is recommended to be used as it produces images of good quality and it does not cause heating and damage to the tissues (Qiu and Schutte, 2007, Švec and Šram, 2011). The image from the laryngoscope is then focused on the image sensor of the system by the focus lens included in the C-mount adapter. The VKG camera head (**Figure 9c**) consists of a beam splitter and two CCD (charge coupled device) image sensors. The beam splitter optically divides the image from the laryngoscope into two paths, each for the two CCD sensors, so that the two-dimensional image (recorded at standard video rate) and kymogram (recorded at high-speed rate) are displayed simultaneously. Of the two CCD image sensors, one is a color area CCD sensor for obtaining two-dimensional image of the vocal folds (for evaluating structural appearance), and the other one is a monochromatic high-speed line-scan CCD sensor for capturing the vocal fold vibrations through a single selected scan-line. By the beam splitter, the scan-line CCD position is fixed on the reflective center of the color CCD, so that the kymogram that is obtained is the vocal fold vibrational behavior from the central line on the two-dimensional structural image of the vocal folds. The camera head is connected to a camera controlling unit (CCU) in a desktop case, which is in turn connected to a video monitor for the display of the VKG images in real time (Švec and Šram, 2011). When using the phase alternating line (PAL) standard, where there are 25 frames per second, each video frame/image is composed of 288 horizontal lines (of 1 pixel height). The spatial resolution of the VKG image (including the standard laryngoscopic image and kymogram together) is about 720 pixels/line, and the total time displayed in the kymogram on the video monitor is 40 millisecond (ms) (Švec and Schutte, 2012, Qiu and Schutte, 2007, Švec and Šram, 2011).

2.7.2.3.2 VKG vibration features and evaluation procedure

When a subject without any voice disorder sustains phonation of a vowel at comfortable pitch and loudness, the following vibratory features should be visible on the kymogram (**Figure 10**) (Šram et al., 2018):

a. Both vocal folds are vibrating
b. Ventricular folds are not vibrating
c. Vibrational amplitudes of both vocal folds are approximately similar
d. Vibrational frequencies of both vocal folds are approximately the same
e. The vibrations are regular
f. The vibrations are free of aberrations
g. The vocal folds touch each other during vibration at the place of maximum vibration amplitude
h. The closed phase takes between ca. 10 and 60 % of period duration at the place of maximum vibration amplitude
i. The shape of lateral peaks (i.e. the turn from opening to closing) is sharp
j. Mucosal waves propagate laterally on the upper vocal fold surface, and
k. No large left-right phase asymmetry is present

The above features are usually obtained with the VKG camera scan line positioned at the place of maximum vibration amplitude that is at the middle of the membranous vocal folds, perpendicular to the glottis. If required the vocal folds may be scanned by moving the camera from anterior to the posterior (A-P scan) along the whole glottal length, in order to obtain the vibrational features at various positions.

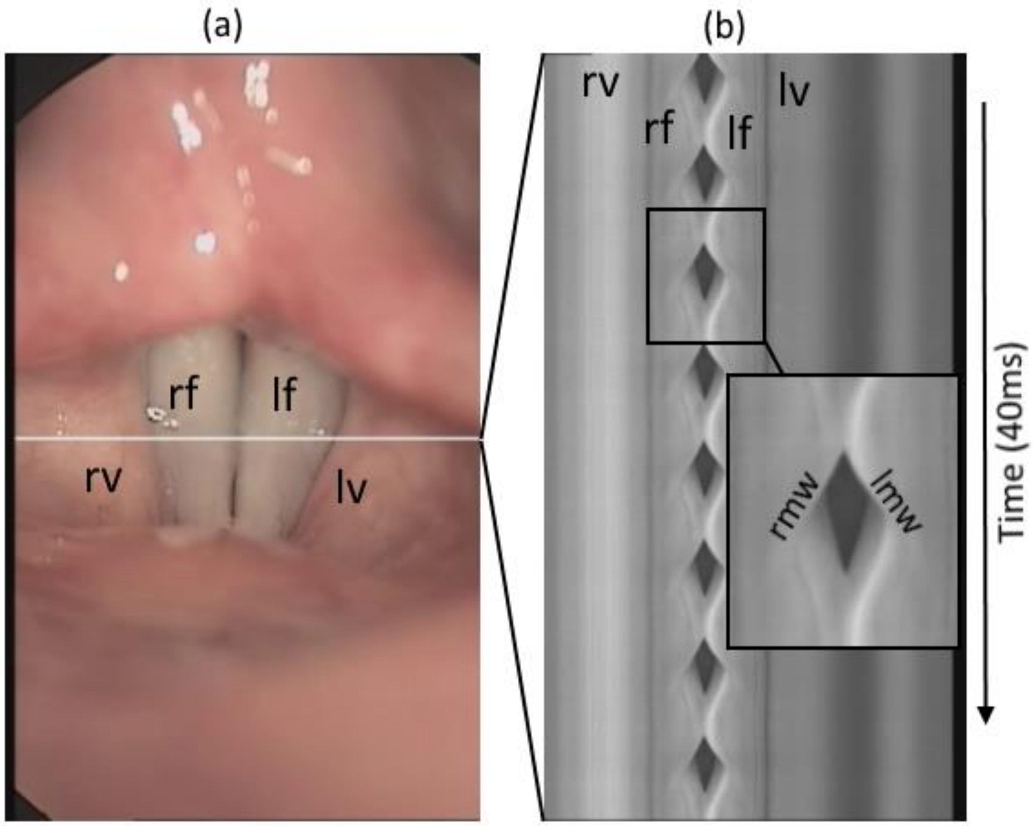


Figure 10: VKG image from a subject with normal vocal fold vibration. (a) Standard laryngoscopic image from the VKG camera during phonation of the right (rf) and left vocal fold (lf). The right (rv) and left ventricular (lv) folds are also visible and not vibrating. The standard laryngoscopic image shows the position of the scan line at the middle of the glottis used for VKG examination. (b) The kymogram (40 ms duration, time running down) shows normal vibrations of vocal folds. The magnified snapshot shows the laterally running right (rmw) and left mucosal waves (lmw) on the right and left vocal folds.

In cases of voice disorders these vibratory features show deviancies from the normal behavior. **Table 5** lists the abnormal VKG features and their possible causes (Šram et al., 2018, Švec and Šram, 2011). **Figure 11** shows VKG images of different types of voice disorders with abnormal vibratory features.

Table 5: Abnormal VKG features and their possible causes (Šram et al., 2018, Švec and Šram, 2011)

Abnormal VKG feature		Possible causes
1	Completely absent vibration of the vocal fold (Figure 11a-rf)	Tumor, scar or excessive vocal fold stiffness.
2	Partly absent vibration of the vocal fold	Stiff vocal fold body
3	Absence of glottal closure (Figure11b)	Problems with vocal fold adduction, lot of breathiness in voice and high-pitched voices.
4	Large cycle-to-cycle variability (Figure 11b and c)	Irregular vocal fold vibrations resulting from left-right vocal fold asymmetry, anterior-posterior vocal fold asymmetry or from excessively low vocal fold tension.
5	Large left-right amplitude differences	Structural abnormalities of left-right vocal folds, such as unilateral vocal fold paralysis.
6	Left-right frequency differences (Figure 11b)	Lack of glottal closure, as in unilateral vocal fold paralysis. The finding usually relates to biphonia or diplophonia.
7	Large left-right phase differences (Figure 11d)	Presence of any mass or tension differences between the right and left folds as in cases of unilateral vocal fold paralysis with good glottal closure (i.e. after medialization surgery), singers complaining of voice problems (perceptually the voice may sound rather normal, but the asymmetry may limit the usable frequency range).
8	Axis shift during closure (Figure 11d)	Differences in tension of right and left vocal folds.
9	Decreased sharpness or rounded lateral peaks (Figures 11a and 11e)	Stiff vocal fold mucosa, due to reduced vertical phase difference of the vocal folds.
10	Absent or reduced mucosal waves on the upper vocal fold surface (Figure 11a-lf, Figure 11e-lf)	Stiffened mucosa on the upper surface of the vocal folds.
11	Sharpened medial peaks	Thinned vocal fold edges
12	Ripple (aberration)	Localized lesion on the vocal fold (nodule, polyp, cyst)
13	Double medial peak (aberration) (Figure 11f, rf-double arrow)	Sulcus vocalis, or a furrow on the medial surface of the vocal folds
14	Medial unsmoothness (aberration)	Defective medial vocal fold shape.
15	Large vibration of surrounding tissues (e.g. ventricular folds or aryepiglottic folds)	Hyperfunction or compensation for glottal insufficiency.
16	Co-vibration of fluids	Too much fluid or mucus interfering with vocal fold vibrations incases of inflammatory disorders such as laryngeal reflux disorder.

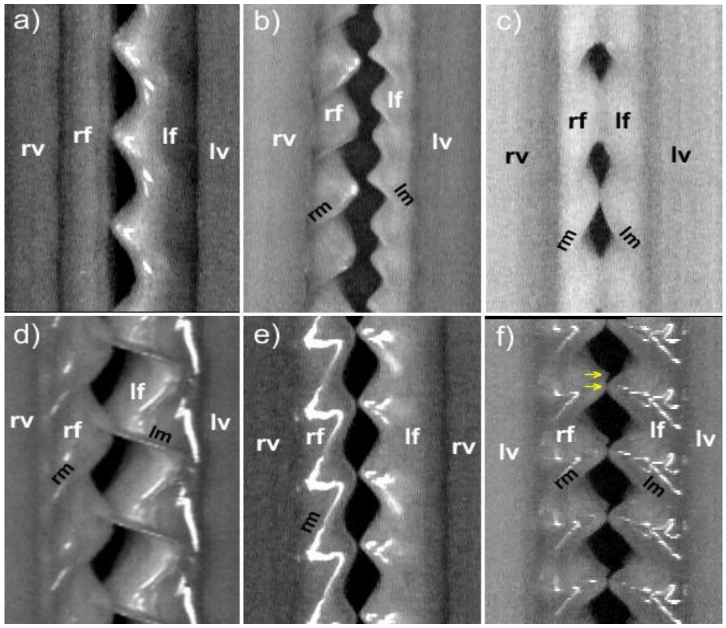


Figure 11: Different types of voice disorders with abnormal VKG features. Total time displayed in the VKG images: 18.4 ms (from top to bottom). Figure taken from Šram et al. (2018).

The VKG vibration features can be evaluated by visual analysis approach based on pictogram representation of these features (Šram et al., 2018). **Table 6** shows the evaluation form with the pictograms for assessing the vocal fold vibrations in clinical practice. A computer assisted software (MS access) which incorporates these visual ratings was also developed, for more detailed evaluation (Švec et al., 2007c) (**Figure 12**).

- ➢ The VKG evaluation form and the computer assisted visual rating software were used in the present thesis for **manuscript I** and **manuscript II** respectively.

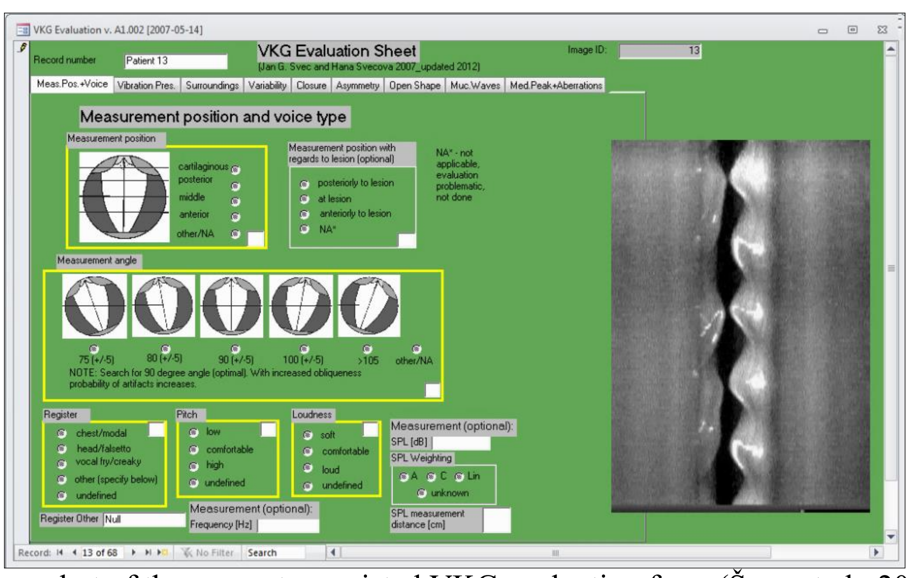
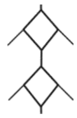
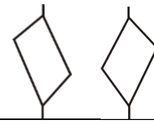
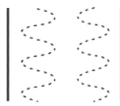
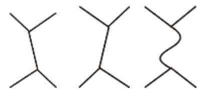
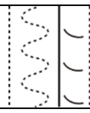
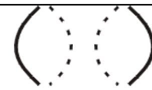
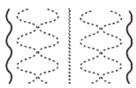
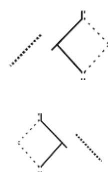
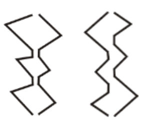
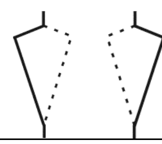
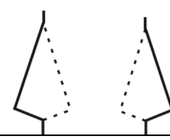
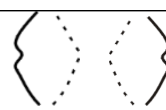
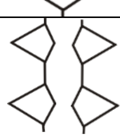
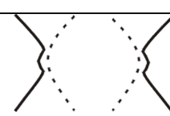
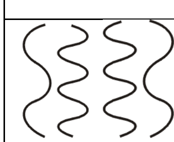


Figure 12: Snap shot of the computer-assisted VKG evaluation form (Švec et al., 2007c).

Table 6: VKG evaluation form

Vibration characteristics		No	Vibration characteristics		No
	Normal vocal fold vibration			Large phase differences	
	Completely absent vibration of the vocal fold			Axis shift	
	Partly absent vibration of the vocal fold			Decreased sharpness of lateral peaks	
	Large vibration of the ventricular fold			Absent or reduced mucosal waves	
	Co-vibration of fluids			Distant mucosal waves	
	Large cycle-to-cycle variability			Opening shorter than closing	
	Absence of glottal closure			Opening longer than closing	
	Short closure duration (CQ< 10%)			Sharpened medial peaks	
	Long closure duration (CQ > 60%)			Ripple	
	Large amplitude differences			Double medial peak	
	Frequency differences			Medial unsmoothness	

2.7.2.3.3 Mucosal waves and lateral peak shape in VKG images

- In the present thesis, we deal with quantification of the shape of the lateral peaks displayed in VKG images (**manuscript II**), and therefore this section elaborates more on this feature.

As mentioned in the beginning of the introduction section (**2.1**), mucosal waves (MWs) have been associated with the time delayed movements between the upper and lower margins of the vocal folds (Titze et al., 1993, Hiroto, 1966), known as the 'vertical phase difference' which facilitates the delivery of airflow energy to the vocal fold tissues (Titze, 1988, Titze, 2000). For the occurrence of MWs, a soft and pliable (or flexible) superficial layer of lamina propria (one of the cover layers of the vocal folds) is necessary (Hirano, 1981a). Presence of MW is an indicator of health and pliability of vocal fold mucosa (Švec et al., 2009). Reduction in the MW amplitude, is an indication of vocal fold lesions and scarring (Bless et al., 1987). Presence of mucosal waves is effectively visualized on the kymograms, and assessed based on (1) vertical phase difference and (2) laterally travelling mucosal waves. The vertical phase difference is indicated by the sharpness of the lateral peak, and the MWs show up as laterally running lines along the upper margin during medial excursion of the vocal folds (Švec et al., 2007b, Švec et al., 2009, Švec and Šram, 2011) as shown in **Figure 13A**. Also the MWs extent (outreach of the mucosal wave to over half of the visible vocal fold) (Švec et al., 2009, Bless et al., 1987) can be interpreted as shown in **Figure 14d**. This extent is reduced, when the vocal folds are excessively stiffened due to vocal fold pathology (Švec et al., 2009, Švec and Schutte, 2012).
The vertical phase difference (VPD) is indicated by the shape of the lateral peak in the kymogram which is observed to vary between sharp and rounded shapes. The lateral peak shapes are formed by the position of the boundary of upper margin of the vocal fold during opening phase, and by the position of the boundary of lower margin during the closing phase of the vocal folds (**Figure 15A**). The glottal edge shifts from the upper to lower margin, at the point of transition between opening and closing phase of the vocal folds. A large VPD renders an abrupt glottal shift between the upper and lower margin, resulting in a sharp lateral peak (**Figure 15A**). In smaller VPD, the glottal edge shift is gradual, resulting in a rounded lateral peak shape (**Figure 15B**). The large VPD is due to increased pliability of mucosal waves and increased vertical thickness of the vocal folds. In contrast, a smaller VPD is a result of increased stiffness of mucosa leading to less efficient transfer of airflow energy to the vocal folds (Švec et al., 2009, Švec et al., 2007b).

The 'vertical phase difference' and the 'shape of the lateral peak' parameters are considered to be diagnostically important for assessing various voice disorders (Švec et al., 2009, Švec et al., 2007b, Shaw and Deliyski, 2008, Švec et al., 2007a, Švec and Šram, 2011). Various authors have tried to measure the vertical phase delay (Titze et al., 1993, Jiang et al., 2008, Shaw and Deliyski, 2008), however very few studies have tried to quantify the shape of the lateral peak parameter (Yamauchi et al., 2016b, Yamauchi et al., 2015). The **manuscript II** in the present thesis investigates the parameters that can be helpful in quantifying the lateral peak shape and correlating it with the visual evaluation approach.

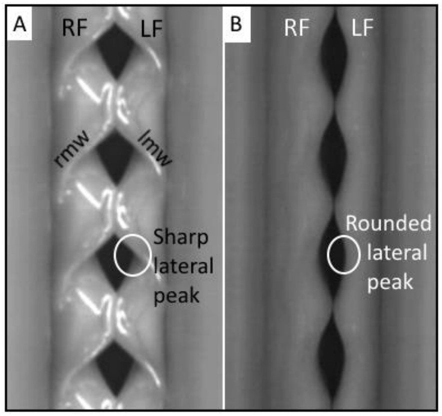


Figure 13: Videokymographic images (with four vibratory cycles) showing (A) sharp lateral peaks (encircled) and laterally running mucosal waves (rmw, lmw) on the right and left vocal fold, respectively; (B) rounded lateral peaks (encircled) with no mucosal waves. RF, LF – right and left vocal fold. The total time displayed here is 17.6 ms (from top to bottom). Figure taken from Kumar et al. (2018), **manuscript II** of present thesis.

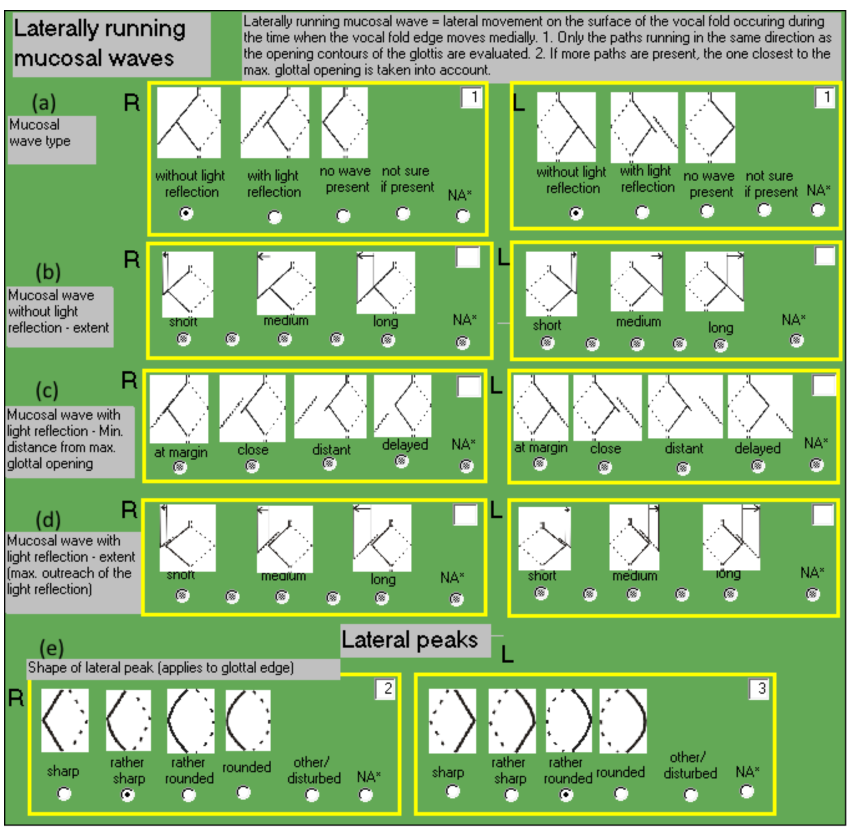


Figure 14: Pictogram showing the laterally running mucosal waves and different lateral peak shapes (Švec et al., 2007c).

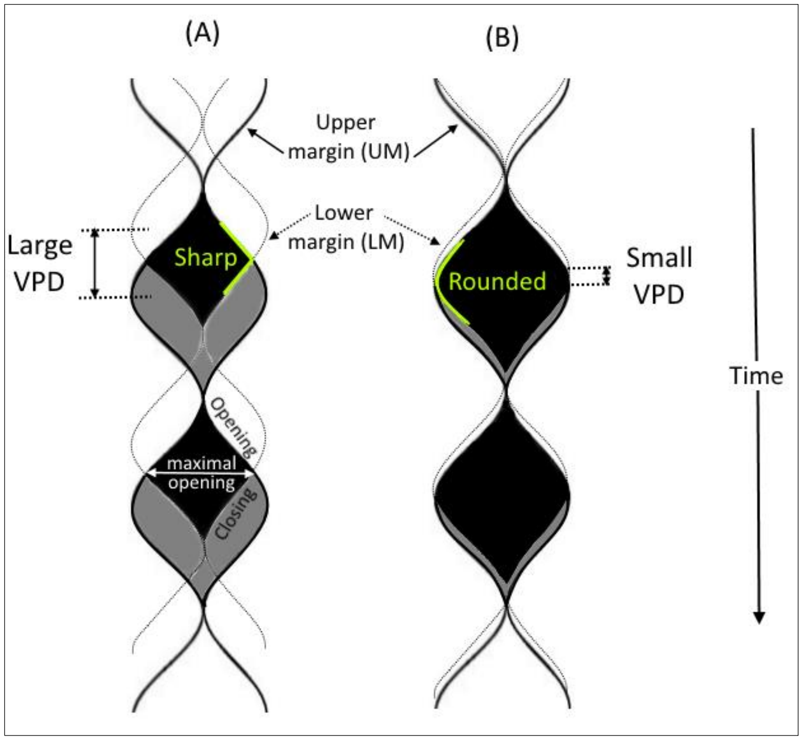


Figure 15: Formation of sharp (A) and rounded (B) lateral peaks in the kymogram. Movements of the lower and upper margins of the vocal folds are indicated by thin-dotted and thick-solid curves, respectively. The vibratory displacement of the lower margin precedes that of the upper margin, thus creating a vertical phase difference between their respective motions. During the opening phase, the motion of the lower margin is invisible indicated by the thin-dotted line; it becomes only visible during the closing phase. A sharp lateral peak (A) is seen when the vertical phase difference is large and a rounded lateral peak (B) is seen when the vertical phase difference is small (indicated in green). LM, lower margin; UM, upper margin; VPD, vertical phase difference. Taken from Kumar et al. (2018), **manuscript II** of present thesis.

2.7.2.3.4 VKG analyzer software

For the quantitative evaluation of VKG features, a computer assisted software (VKG analyzer software) has been developed which works based on digital image processing technique (Novozámský et al., 2015). This software complements the visual evaluation approach of VKG images (VKG evaluation form, **Table 6**), which is time consuming and clinician-dependent. The VKG analyzer software is an automatic software tool that involves (1) data processing, (2) extraction of vocal fold characteristics and (3) detection and extraction of the important VKG features. This tool provides a quantitative data for features such as: number of cycles, amplitude, phase and frequency difference, cycle-to-cycle variability in left and right vocal fold, axis shift, skewing of left and right vocal fold (related to closing and opening phases of vocal fold), and triangularity of left and right fold (related to shape of lateral peak parameter). The quantification of these features is based on accurate detection and extraction of contours defining the glottal edge boundary of the left and right vocal folds (indicated by blue lines in the magnified kymogram in **Figure 16**). These extracted contours can be saved in text files as set of data defining the glottal edges of both vocal folds along with their time instances.

- The glottal edge contours obtained from the VKG analyzer software have been used for further processing using MATLAB script in the present thesis in **manuscript II**, for quantification of shape of the lateral peak parameter.

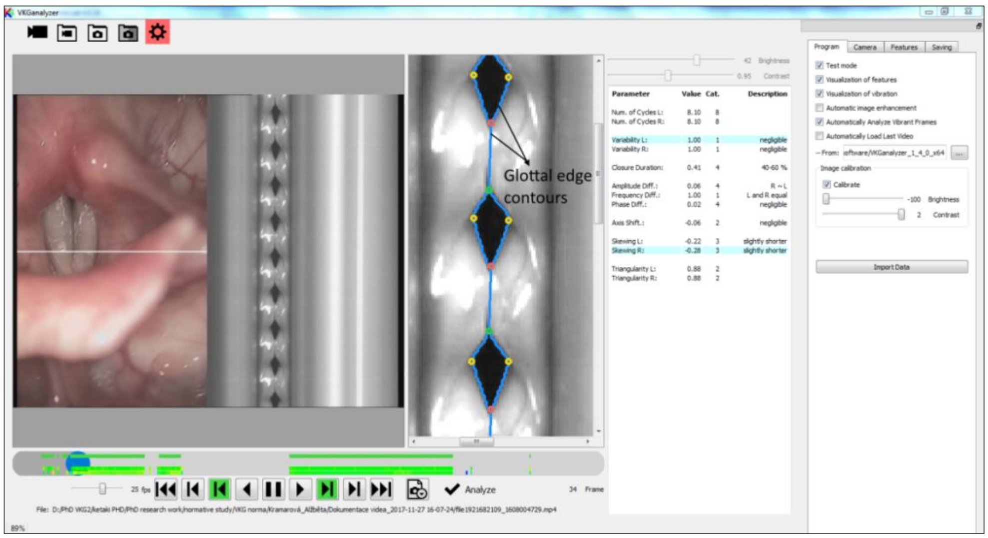


Figure 16: Screenshot of the VKG analyzer software (Novozámský et al., 2015) showing the detected glottal edge contours on the right and left vocal folds, indicated by blue lines in the magnified kymogram.

2.7.2.4 Clinical value of stroboscopy, HSV and VKG

Stroboscopy is considered to be the current “golden standard” method for diagnosis and treatment of various voice disorders. Clinical value of stroboscopy has been explored in various studies. These studies compare the value of videostrobocopy in assessing and treating various voice disorders with the traditional visualization methods such as indirect mirror and fiberoptic examination of larynx (Woo et al., 1991), subjective impressions and laryngoscopy under continuous light (Sataloff et al., 1991), indirect laryngoscopy (Casiano et al., 1992), case history with physical examination and flexible laryngoscopy (Paul et al., 2013) and continuous light rigid laryngoscopy (Printza et al., 2012). These authors report stroboscopy to have confirmed or changed the initial diagnosis and treatment decisions made based on these traditional visualization techniques, thus proving its clinical usefulness. **Figure 17** summarizes the usefulness of stroboscopy reported by the above mentioned authors. Stroboscopy has been helpful in diagnosing general hoarseness (Sulica, 2011), specific pathological conditions such as vocal fold scar and other organic lesions (Mehta and Hillman, 2012, Rosen et al., 2012).

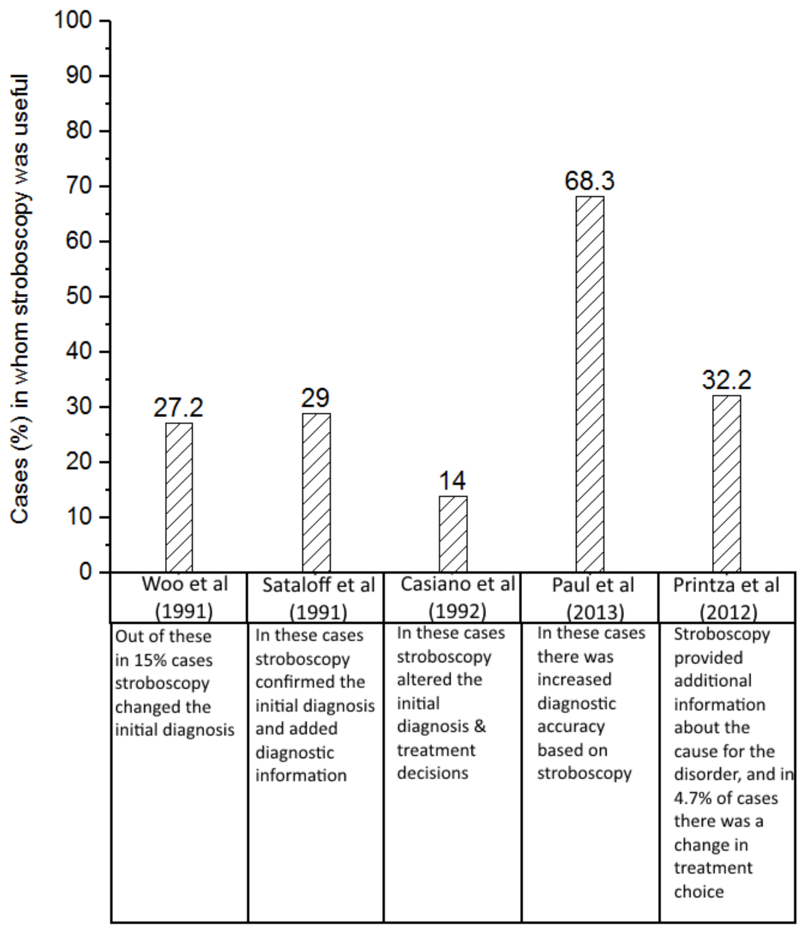


Figure 17: Findings of the studies reporting clinical usefulness of stroboscopy in their patients.

Although videostroboscopy has been a very useful diagnostic tool, it is confined to only periodic vibrations of the vocal folds, thus hampering the accurate diagnosis of hoarse voice quality. With the development of high-speed imaging technique with high sampling rate, it is now possible to analyze cycle to cycle variations in the vocal fold vibrations. The clinical value of HSV has also been explored in many studies. **Table 7** summarizes the findings of studies reporting the usefulness of HSV over videostroboscopy for diagnosis of voice disorders. HSV has been reported to be useful in assessing vibratory functions in moderate to severe voice disorders, that are not interpretable on videostroboscopy (Patel et al., 2008). HSV has also been used as an alternative to videostroboscopy especially for the diagnosis of functional voice disorders, since it allows assessing phonation onset, irregularities and all aperiodic patterns of vocal fold vibration (Braunschweig et al., 2008). It is also advantageous in examining voice production mechanism in individuals with postsurgical early glottic cancer (Mehta et al., 2010b) and analyze vibratory changes quantitatively after vocal fold surgery (Chen et al., 2016).

Table 7: Findings of the studies reporting clinical usefulness of HSV for diagnosing voice disorders

Studies	Aim	Findings
Patel et al. (2008)	Investigated the clinical value of HSV compared to videostroboscopy across three disorder groups (epithelial, subepithelial and neuromuscular disorders)	About 63 % participants' voice recordings (100 % severe voice disorders and 64 % moderate voice disorders) were non-interpretable for assessment of vibratory functions on videostroboscopy, compared to HSV, which helped assess 100 % of the participants. Neuromuscular group (paralysis, spasmodic dysphonia, atrophy or bowing, primary muscle tension dysphonia) was most difficult to be diagnosed on videostroboscopy followed by epithelial (caricinoma, reflux, laryngitis) and lastly subepithelial (nodules, polyp, scar, sulcus vocalis) voice disorders.
Kendall (2009)	Compared usefulness of HSV over videostroboscopy in healthy subjects	There was no difference between the two techniques for any of the vibratory features except for periodicity. Aperiodic vibratory feature was noted in 30 % cases on videostroboscopy and only in 4 % on HSV leading to appropriate diagnosis.
Mendelsohn et al. (2013)	Evaluated the diagnostic role of HSV over videostroboscopy for various voice disorders	There was no improvement in the diagnostic accuracy between the two techniques, however presbyphonia was better diagnosed on HSV.
Zacharias et al. (2018)	Investigated the usefulness of HSV for clinical voice assessment over videostroboscopy for various voice disorders	There were changes in the ratings of vibratory features between videostroboscopy to HSV: In 74 % cases ratings of severely decreased mucosal waves on videostroboscopy changed to moderately decreased mucosal wave on HSV. There was also change in the rating of amplitude of vibration feature (53 %), refinements in ability to describe glottal closure patterns (36 %), phase symmetry either decreased or increased (21 %). In 7 % cases there was change in initial diagnosis and intervention recommendations after HSV examination.

Although HSV has been reported to be useful in diagnosing various voice disorders, due to its limitations pertaining to more time consumption and an advanced offline objective analysis, it is rather difficult to be implemented as a diagnostic tool regularly in a busy clinic where there is relatively large patient load. Therefore, VKG which allows for analysis of vocal fold vibratory behavior in real time is a more beneficial tool that can be used in clinics. It allows for immediate clinical feedback, that is, the examiner is able to visualize the VKG images simultaneously during the examination and discuss about the problems with the patient. Although VKG has been reported to be helpful in better recognizing some voice disorders (Švec et al., 2007b, Piazza et al., 2012, Chodara et al., 2012), its diagnostic value and clinical relevance for treatment recommendations has not been formally evaluated. Although in some clinics VKG is used routinely, many clinicians are unaware of its clinical usefulness. Thus in the present thesis, **manuscript I** evaluates the clinical value of VKG in diagnosing various voice disorders.

2.8 Acoustic evaluation of voice

Recording a voice signal using a microphone is the most non-invasive procedure of evaluating the vocal performance of an individual. Acoustic analysis of voice has its application in voice screening and diagnostics, and monitoring the treatment effectiveness (Hirano et al., 1988).

ELS (Dejonckere et al., 2001) in their basic protocol recommended the use of perturbation measures (percept jitter and percent shimmer) as robust measures of voice quality. Perturbation refers to the random disturbances in periodic motion. Pitch perturbation and amplitude perturbations are related to perceptual parameters of vocal pitch, loudness and quality.

Pitch perturbation refers to the cycle-to-cycle variations in the fundamental frequency of voice signal known as fundamental frequency perturbation or the *'vocal jitter'*. Jitter may be computed as jitter percentage (%). The jitter percentage is also known as jitter factor. It is the mean difference between the frequencies (Hz) of adjacent cycles divided by mean frequency of voice signal. It is calculated based on the following formula (Awan, 2001, Kent and Ball, 2000):

$$\frac{\frac{1}{n-1}\left[\sum_{i=1}^{n-1}|F_i - F_{i+1}|\right]}{\frac{1}{n}\sum_{i=1}^{n} F_i} \times 100$$

Where n is the number of periods in a sample, i=1, 2, 3...n; F_i is the frequency of the i^{th} cycle in Hz. The MDVP (Multi-Dimensional Voice Program, Model 5105) provides jitter threshold/normative value of 1.04 % in its operating manual by Kay Elemetrics (2008).

Amplitude perturbation, also called '*vocal shimmer'*, is defined as cycle-to-cycle variations in the amplitudes of adjacent pitch pulses, expressed in decibels (dB), but can also be expressed in percentages (%). Vocal Shimmer represented in (%) is based on the following formula (Kent and Ball, 2000):

$$\frac{\frac{1}{n-1}\left[\sum_{i=1}^{n-1}|(A_i - A_{i+1})|\right]}{\frac{1}{n}\sum_{i=1}^{n} A_i} \times 100$$

Where n is the number of cycles, i=1, 2, 3...n; A_i is the peak to peak amplitude of the i^{th} cycle in dB. The MDVP (Multi-Dimensional Voice Program, Model 5105) provides shimmer threshold of 3.810 % in its operating manual by Kay Elemetrics (2008).

These measures are routinely being used in clinical and research setups. Extensive research has been carried out for investigation of application of these perturbation measures in assessing and finding the treatment effectiveness of voice disorders, and correlating it with subjective voice ratings (Fex et al., 1994, Lopes et al., 2017, Niedzielska, 2001, Hirano et al., 1988, Kasuya et al., 1986, Hammarberg et al., 1980). However one of the major technical limitations of jitter and shimmer measure is that it requires precise determination of the fundamental frequency (fo) and amplitude, which is technically not easy. Even small errors in this determination can lead to large effects on the perturbation measures (Yiu, 1999, Mehta et al., 2011). Therefore, these two measures are considered not very reliable in cases of highly aperiodic voice signals, because in these signals the fo varies across time, and detection of accurate fo is difficult through the use of pitch detection algorithms used in time domain (Titze, 1995, Titze and Liang, 1993). This is especially problematic in cases of moderate to severe dysphonic patients, where the voice signals are less periodic and accurate extraction of fo is difficult (Heman-Ackah et al., 2003). Recently, the cepstral analysis measure has emerged as a better acoustic measure that quantifies the voice signal and extracts the fo accurately in frequency domain. Also this measure has been considered to be a reliable correlate of overall severity of dysphonia (Heman-Ackah et al., 2003) and has also been recommended to be used in the assessment protocol provided by ASHA (Patel et al., 2018).

- The next section will deal with details on cepstral analysis of voice, which is of particular interest in **manuscript III** of this thesis.

2.8.1 Cepstral analysis of voice

The *'cepstrum analysis'* is a more advantageous method than the traditional acoustic (jitter and shimmer) measures for extracting the fundamental frequency of voice. This measure was originally described by Noll (1964), and then explored by Hillenbrand and colleagues who developed a measure of breathy voice quality called "the cepstral peak prominence" (Hillenbrand et al., 1994, Hillenbrand and Houde, 1996).

2.8.1.1 The cepstrum concept

To understand what a cepstrum is and how it is created, refer to **Figure 18,** which illustrates this concept. First the acoustic waveform of a voice signal (for example phonation of the vowel [a:]) is subjected to fast Fourier transform, resulting in a *frequency spectrum*, which is the action of transitioning from time-domain to frequency-domain of the signal. The spectral components are then squared to generate a power spectrum, after which the logarithm of each of the power spectrum term is taken. The obtained log power spectrum is again subjected to a new Fourier transform, considering the spectrum itself to be a time-domain waveform, thus creating the *cepstrum* in quefrency-domain (or 1/frequency-domain). The 'ceps' in cepstrum is the inverse of the term 'spec' in spectrum. The horizontal axis of the cepstrum (**Figure 18**) is designated as 'quefrency' to differentiate it from 'frequency' of the spectrum, measured in milliseconds (ms). In the cepstrum, the high-quefrency sharp dominant peak represents the cepstral peak. The peaks beyond the cepstral peak are the 'rahmonics' which is the equivalent term for harmonics in the spectrum, and the lower-quefrency peaks below the cepstral peak are related to the resonance of cavities above the vocal folds. The fundamental period of the voice signal is represented by the sharp cepstral peak, and the fundamental frequency is the reciprocal of the quefrency at this peak (Baken and Orlikoff, 2000). The cepstrum is thus defined as a log power spectrum of a log power spectrum (Hillenbrand and Houde, 1996).

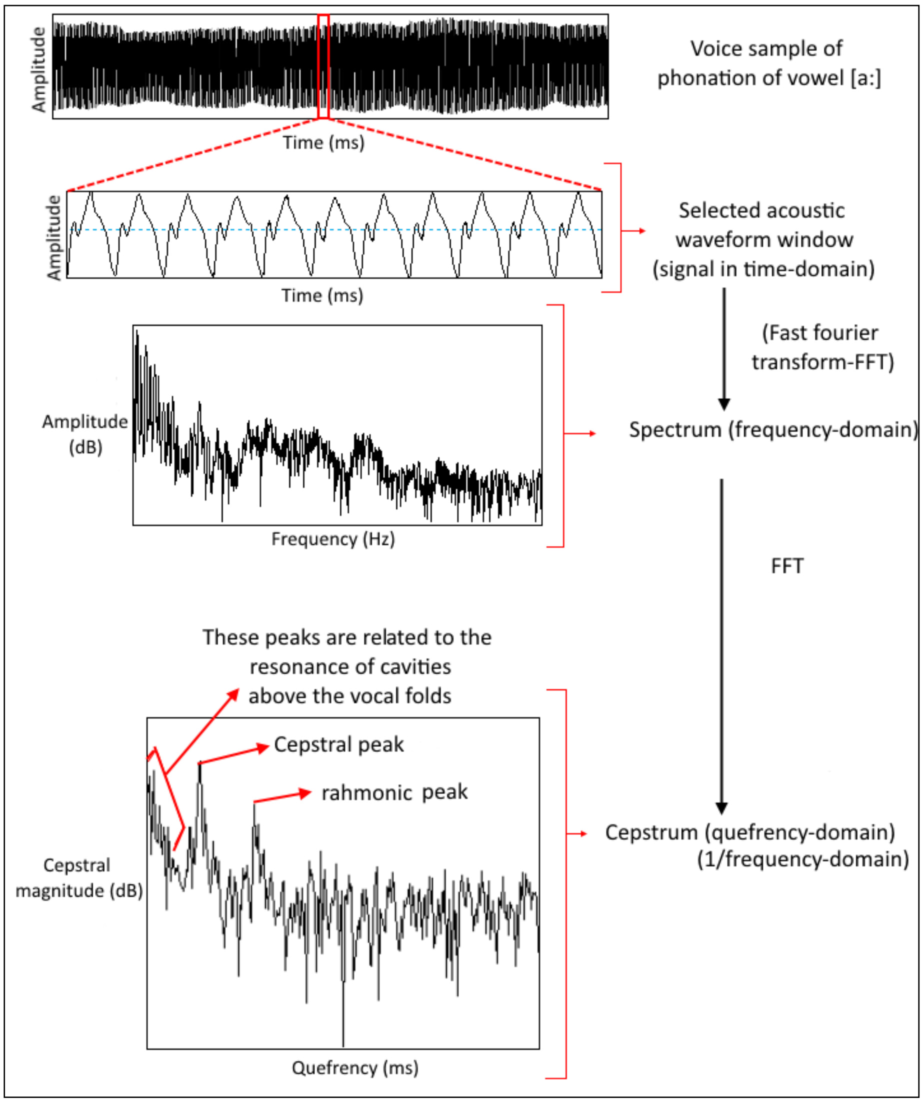


Figure 18: Flowchart depicting the steps involved in creating a cepstrum. Figure modified from McDonald et al. (2011).

2.8.1.2 The cepstral peak prominence (CPP)

Hillenbrand and colleagues (Hillenbrand et al., 1994, Hillenbrand and Houde, 1996) explored the use of cepstral analysis in evaluating breathiness of voice. They developed a metric called *Cepstral Peak Prominence (CPP).* The sharp cepstral peak (corresponding to the fundamental period of the voice signal) mentioned in the previous section, has the highest amplitude. When a linear regression line, representing the average sound energy is drawn through the cepstrum (solid black line shown in **Figure 19**), the distance between the regression line and the cepstral peak (red dotted line in **Figure 19**) is termed as the CPP. In other words, CPP is the difference in amplitude between the cepstral peak occurring within the boundaries of the expected phonational quefrencies and the corresponding value on the regression line fitted on the cepstrum, measured in decibels) (Hillenbrand et al., 1994, Heman-Ackah et al., 2003).

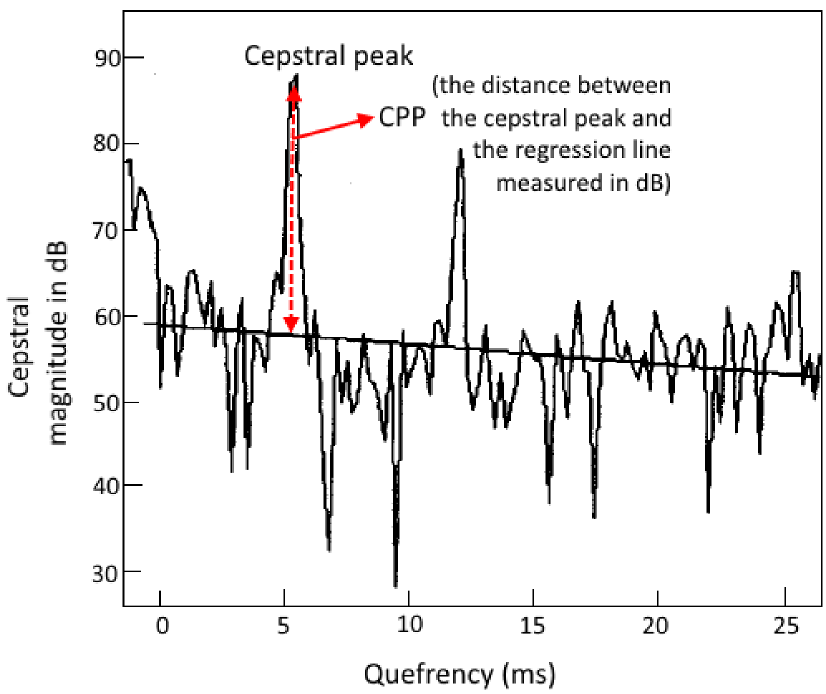


Figure 19: Image showing the regression line fitted on the cepstrum (nearly horizontal black solid line), and the cepstral peak prominence (CPP) (vertical red dotted line) shown as the distance between the cepstral peak value and the regression line value measured in decibels (dB). Figure modified from Hillenbrand and Houde (1996).

The idea behind developing the CPP measure (Hillenbrand et al., 1994) is that a highly periodic signal (as in healthy or non-breathy voice) (**Figure 20a**) presents with a well-defined harmonic structure resulting in a more prominent cepstral peak. However, in a less periodic signal (as in breathy or dysphonic voice) (**Figure 20b),** the cepstral peak is not very prominent due to presence of cepstral background noise. Since CPP is a measure of harmonic organization (or periodicity) over the “noisiness” in the voice signal, a breathy voice tends to have smaller CPP value than non-breathy voice (**Figure 20 (a, b)**). Hillenbrand et al. (1994) found strong negative correlation between the CPP and perceptual breathiness rating (Pearson's correlation coefficient r = -0.9), indicating that the CPP reduced with increase in breathiness perception of the voice signal.

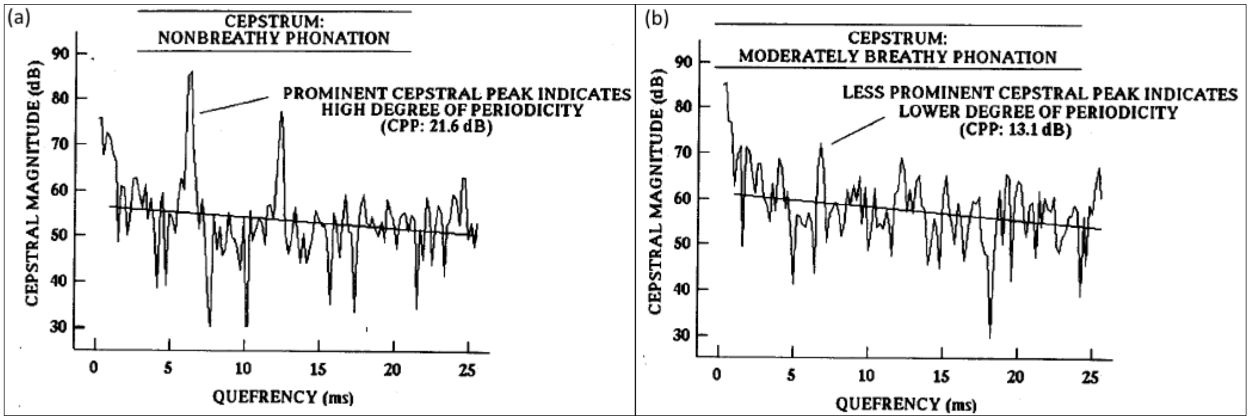


Figure 20: Image showing (a) cepstrum of non-breathy phonation and (b) moderately breathy phonation. Cepstral peak is prominent in non-breathy and not so prominent in breathy phonation. Figure modified from Hillenbrand et al. (1994).

2.8.1.3 Smoothed cepstral peak prominence (CPPS)

The CPP measure was initially developed to analyze sustained vowels (Hillenbrand et al., 1994). Later on a modification of this metric called the *Smoothed cepstral peak prominence (CPPS)* was developed for greater prediction accuracy particularly in speech signals (Hillenbrand and Houde, 1996). This metric is called smoothed CPP because the individual cepstra are smoothed across time and quefrency domains.

The time and quefrency smoothing (or averaging) procedure is illustrated in **Figure 21**. In the first step, individual unsmoothed cepstra are obtained from the spectra derived from the waveform of the voice signal (**Figure 21, step 1**). In step 2 the individual cepstral frames are smoothed across time. The time smoothing can be explained by considering an example of smoothing window of five frames (**Figure 21, step 2**). Here the smoothed output for a given frame would have average of the current frame with two previous frames and the two subsequent frames. In step 3, a running average of cepstral magnitude is calculated across quefrencies, which can be explained using an example of five-bin averaging window (**Figure 21, step 3**). Here each quefrency component is replaced by the average of the current bin with the two adjacent bins of lower quefrency and the two adjacent bins of higher quefrency.
Following the time and quefrency smoothing, the smoothed cepstral peak is extracted and the cepstral peak prominence is calculated as smoothed cepstral peak prominence (CPPS). Similar to CPP, the CPPS is also calculated as the difference between the smoothed cepstral peak value and the value on the regression line fitted on the smoothed cepstrum as shown in **Figure 22**.

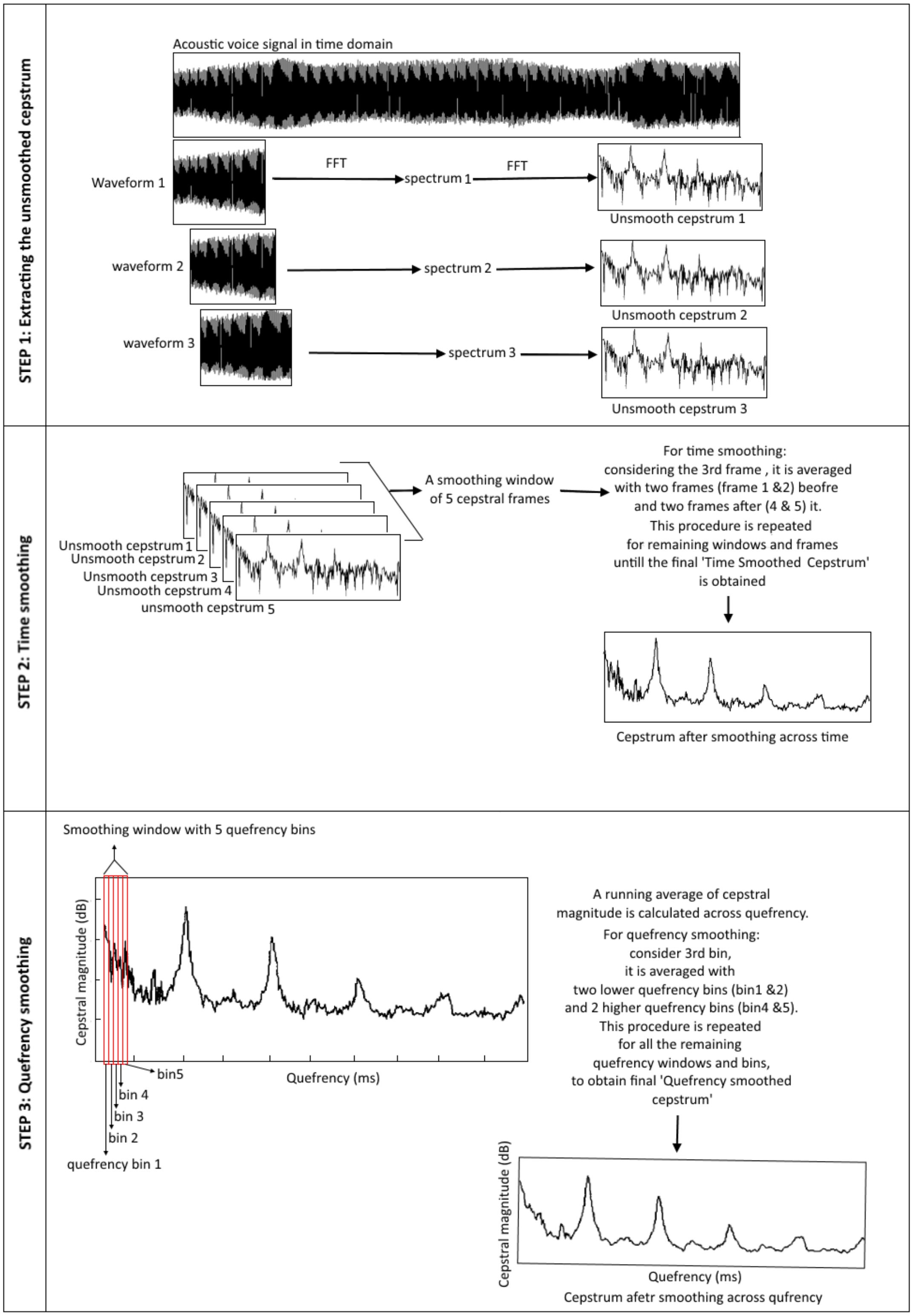


Figure 21: Image illustrating the procedure of extracting the unsmoothed cepstrum in step 1, followed by time and quefrency smoothing in step 2 and 3 respectively. Figure modified from Hillenbrand and Houde (1996).

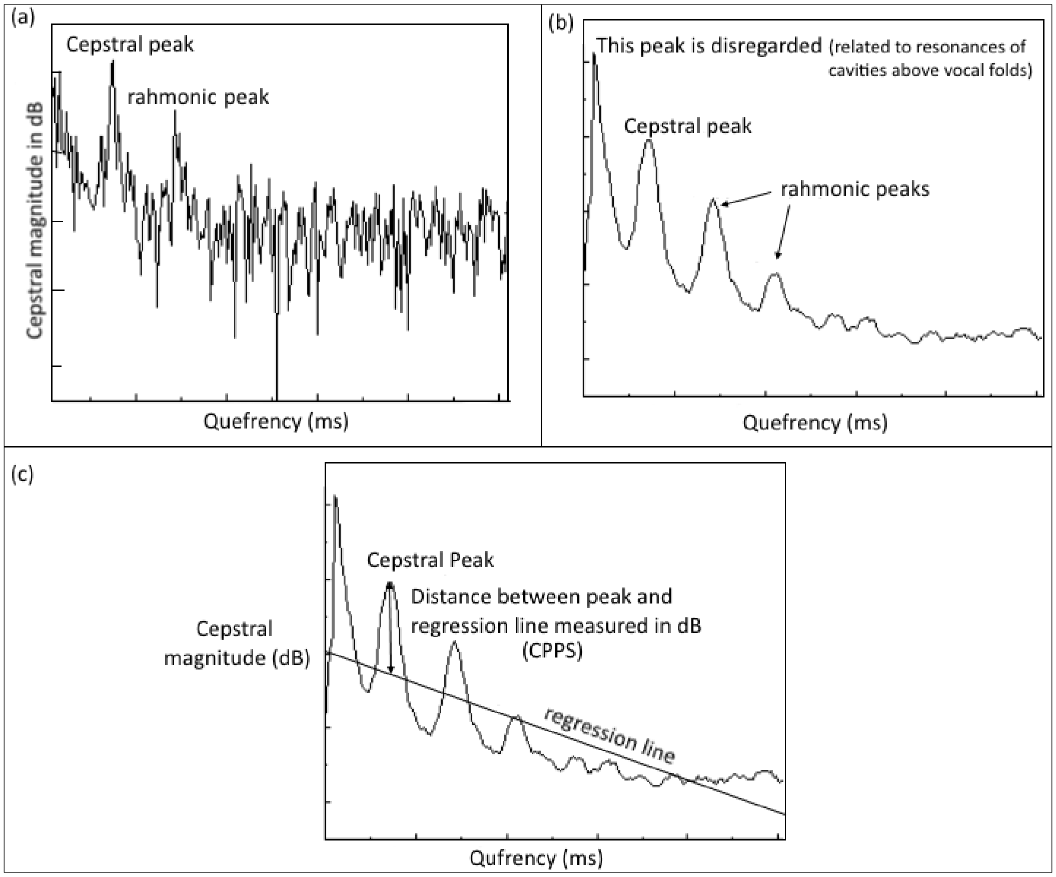


Figure 22: Image showing the (a) unsmoothed cepstrum, (b) smoothed cepstrum and (c) smoothed cepstrum with regression line fit. CPPS is the smoothed cepstral peak prominence calculated as the distance between the smoothed cepstral peak and regression line fitted on the cepstrum). Figure modified from McDonald et al. (2011).

In the Hillenbrand's study, since a continuous speech sample had a varying fundamental frequency over time, a 10 frame (20 ms) time smoothing window, followed by a 10 bin quefrency smoothing window was chosen (Hillenbrand and Houde, 1996). In his study, it was reported that for both vowel and continuous speech samples, CPPS correlated strongly and provided accurate predictions of perceptual breathiness rating (Pearson's correlation coefficient r = -0.93 for vowel and r = -0.92 for speech). Also CPP had good correlation with breathiness rating for both the samples (r = -0.89 for vowel and r = -0.88 for speech).

2.8.1.4 Extraction of CPP and CPPS using different software packages

Extraction of CPP and CPPS has been implemented in various software packages. Hillenbrand was the first to develop the *SpeechTool* software for the extraction of CPP and CPPS measures (Hillenbrand, 2006, Hillenbrand and Houde, 1996, Hillenbrand et al., 1994). Later, these measures have also been implemented in other software packages such as *Computerized Speech*

Lab (CSL, Kay Pentax, Lincoln Park, NJ) and the freely available *Praat* software (Maryn and Weenink, 2015). Unfortunately, different software packages make use of different algorithms for the calculation of the cepstral values, limiting the comparison of these values across studies using different software. For example, Heman-Ackah et al. (2014) have reported a mean CPPS value of 4.77±0.97 dB in 30 normal voices on a running speech/reading sample, and also Hasanvand et al. (2017) reported a mean CPPS value of 5.411±1.204 dB in 100 normal voices on a reading sample. Both these studies made use of Hillenbrand's algorithm/*SpeechTool* software for cepstral analysis. However, the cepstral values extracted using the *Praa*t software, show largely different values. For example, Latoszek et al. (2018) reported a mean CPPS value of 11.92±2.15 dB in individuals with perceptually non-dysphonic voices using the Acoustic Voice Quality Index (AVQI) based CPPS setup (*Praat* version 5.3.57). Therefore comparing the CPPS values obtained from *SpeechTool* software and *Praat* software, there is an approximate difference of 5 to 7 dB in the values between the two software packages. Also in a study by Kim et al. (2017) the CPP values for the same voice sample extracted from *Praat*, *SpeechTool* software and Analysis of Dysphonia in Speech and Voice (ADSV) software largely differed, but showed a strong correlation (Pearson's r >0.95) among the measures. This relationship was true for both vowel and speech samples. This strong correlation indicated that CPP value could differentiate a wide range of voice signal periodicity in a similar fashion irrespective of the software it was extracted from. Similar correlations were obtained for CPP measures extracted from ADSV and *Praat* software reported by Watts et al. (2017). Madill et al. (2018) also found good correlations among the cepstral measures derived from ADSV, *SpeechTool*, and VoiceSauce software. However, the authors report different threshold values for detecting dysphonia severity across different software programs.

Differences in the cepstral values are not only seen between different software packages, but discrepancies also exist within different versions of the same software. An example of such software is *Praat* which specifies different default time and quefrency averaging window settings across different versions. Maryn and Weenink (2015) have recommended using a time averaging window of 0.01s and quefrency averaging of 0.001s for obtaining CPPS in their AVQI-based *Praat* setup. However, these settings are different from the standard/default settings of *Praat* for extracting CPPS values. Also the standard settings change with different *Praat* versions. **Table 8** shows the procedure for extraction of CPPS in *Praat* software, using the standard settings of *Praat* and the settings recommended by Maryn and Weenink (2015). **Table 9** shows the CPPS values of a continuous speech sample for a female subject for these two settings and the comparison of the values across three different *Praat* versions (version number 5.3.56, 5.4.04, and the latest *Praat* version 6.0.43 updated on 25/10/2018) is also shown.

Table 8: The steps and parameter settings in the *Praat* software (*version* 5.4.05) for extraction of CPPS values for voice samples (vowel or continuous speech)

Step 1) Select the vowel or speech sample **Step 2)** Go to 'Analyse periodicity' and click on to 'To Power cepstrogram' in the *Praat* Objects window. **Step 3**) Use the following settings for generating the power cepstrogram:		
Parameter setting	**Standard settings of *Praat***	**Settings recommended by Maryn and Weenink (2015)**
Pitch floor (Hz)	60	60
Time step (s)	0.002	0.002
Maximum frequency (Hz)	5000	5000
Pre-emphasis from (Hz)	50	50
Step 4) On selecting the newly generated 'powercepstrogram' click on to 'Query'and select 'Get CPPS' from the menu, and use the following settings:		
Parameter setting	**Standard settings of *Praat***	**Settings recommended by Maryn and Weenink (2015)**
Select subtract tilt before smoothing	Yes	No
Time averaging window (s)	0.001 (*subject to change with different Praat versions*)	0.01
Quefrency averaging window (s)	0.00005 (*subject to change with different Praat versions*)	0.001
Peak search pitch range (Hz)	60-330	60-330
Peak search tolerance (0-1)	0.05	0.05
Interpolation	Parabolic	Parabolic
Tilt line quefrency range (s)	0.001-0.0 (=end)	0.001-0.0 (=end)
Line type	Exponential decay	Straight
Fit method	Robust	Robust

Table 9: CPPS values from continuous speech sample of a female subject for different Praat versions and different time and quefrency window settings

***Praat* versions**	**Settings**	**Time averaging window (s)**	**Quefrency averaging window (s)**	**CPPS value in dB**
5.3.56	Maryn and Weenink (2015)	0.01	0.001	5.58
5.4.04				5.32
6.0.43 (updated on 25/10/2018)				5.32
5.3.56	Standard settings	0.001	0.00005	15.17
5.4.04		0.001	0.00005	14.45
6.0.43 (updated on 25/10/2018)		0.02	0.0005	5.43

From **Tables 8 and 9,** considering the CPPS settings recommended by Maryn and Weenink (2015), the time averaging window of 0.01 s and time step of 0.002 s imply that the time averaging is done over 5 successive cepstra. Also, the quefrency averaging window of 0.001 s and the quefrency step of 0.0001 s (derived from the maximum frequency of 5000 Hz) reveal that the quefrency averaging is done over 10 quefrency bins. Similarly, considering the standard settings for the latest *Praat* version 6.0.43, the time averaging window of 0.02 s with time step of 0.002 s and quefrency averaging window of 0.0005 s with quefrency step of 0.0001 s, results in smoothing over 10 successive cepstra and 5 quefrency bins. This suggests that the cepstral values extracted from recommednations by Maryn and Weenink (2015) for versions 5.3.56, 5.4.04 and 6.0.43, and the standard settings for the version 6.0.43 are truly the smoothed CPP values. However, in the standard settings for versions 5.3.56 and 5.4.04, the time averaging window of 0.001 s and quefrency averaging window of 0.00005 s are so short that effectively no smoothing takes place. This indicates that the values 15.17 dB and 14.47 dB (**Table 9**) are not be the smoothed CPP (i.e. CPPS) values but rather the non-smoothed (CPP) ones (despite of being called CPPS). It is therefore necessary to cross-check priorly whether the time and quefrency averaging window settings in *Praat* are accurately providing the smothed or non-smoothed CPP values.

Furthermore, taking a closer look at the cepstral values for versions 5.3.56 and 5.4.04 in **Table 9**, the cepstral values are different inspite of using the same settings, be it for the Maryn and Weenink's settings (5.58 and 5.32 dB) or the standard settings (15.17 and 14.45 dB). The reason for such differences in the cepstral values is not clear and calls for more carefulness in using freely available software packages. Moreover, the standard *Praat* settings for version 6.0.43 have changed resulting in a completely different cepstral value compared to the values obtained by the standard settings of previous versions 5.3.56 and 5.4.04. It is therefore important not to rely blindly on standard settings but to choose appropriate settings for the extraction of CPP and CPPS values in order to have effective comparison across studies and better reproducibility of the results. This is especially important for speech pathologists or clinicians who are not very technically experienced, and may use different software packages blindly without adjusting or checking the settings. In the present thesis, **manuscript III** elaborates on the appropriate settings for extracting CPP and CPPS from *Praat* software for better comparisons of results across studies.

2.8.1.5 Applications of CPP and CPPS for assessing voice

Ever since Hillenbrand and his colleagues applied CPP and CPPS measures to breathiness evaluation, researchers have explored the use of cepstral analysis to evaluation of dysphonia. Heman-Ackah et al. (2002) reported that the CPPS measure for both vowel and continuous speech samples were best predictors of overall dysphonia, and was also considered as a more robust measure compared to traditional acoustic measures such as noise to harmonic ratio (NHR) and other shimmer and jitter related measures. CPPS measure (extracted from the CSL software which uses Hillenbrand's algorithm) was found to also predict the severity of dysphonia, with a value above 10 dB to be normal for vowel phonation and a value above 5 dB to be normal for continuous speech samples (Heman-Ackah et al., 2003). Maryn et al. (2009) found the cepstral measures, particularly the CPPS measure, to be the most robust single correlate of overall voice quality based on their meta-analysis results. The authors argue that the perception of dysphonia increases when the periodicity of the voice signal attenuates. Cepstral peaks, which are related to the prominence of the fundamental frequency (i.e., periodicity) of a voice signal, may be therefore considered robust for assessing overall voice quality. **Figure 23** shows how the CPPS is affected in moderate and severely dysphonic voices compared to normal voice signal, for a vowel phonation sample.

CPP and CPPS measures have been found to correlate strongly with perceptual evaluations of voice (Heman-Ackah et al., 2002, Awan et al., 2010). CPP measures have been helpful in differentiating between various dysphonia types. It has been reported that CPP values are higher for pressed and normal (modal) phonation compared to breathy type of phonations (Shue et al., 2010). These findings have been attributed to larger open quotient values of glottal waveform during breathy phonations which lead to increased spectral noise (Shue et al., 2010). Wolfe and Martin (1997) applied four parameter model, including CPP measure to classify dysphonic patients into breathy, hoarse and strained voice quality. The CPP values were lower for hoarse and breathy voice compared to strained voice type. The CPP measure has been found to be useful for differentiating hypofunctional from normal voice (Watts and Awan, 2011), in assessing voice quality in various voice disorders (Zieger et al., 1995), vocal nodules (Kumar et al., 2010) and unilateral vocal fold paralysis (Balasubramanium et al., 2011). CPPS has been found to have a high predictive value of voice disorders and therefore has been recommended for voice screening purposes (Sauder et al., 2017).

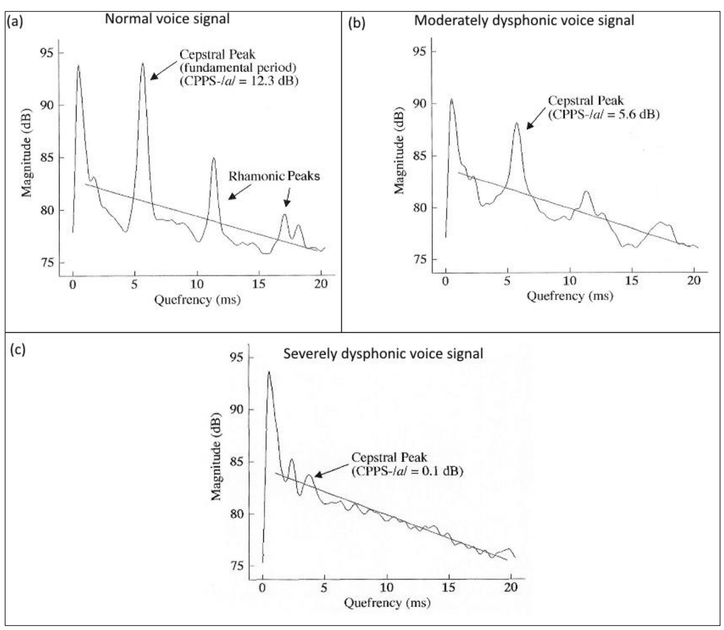


Figure 23: Differences in the CPPS values of vowel phonation for (a) normal, (b) moderately dysphonic and (c) severely dysphonic voice signals. Figure modified from Heman-Ackah et al. (2003).

- In the present thesis, the use of CPP and CPPS measures has been explored in assessing voice of teachers using the *Praat* software in **manuscript III**.

2.8.2 Voice sound pressure level

The measurement of vocal loudness (or voice Sound Pressure Level - SPL) has also been an important acoustic measurement considered as part of voice assessment protocol (Patel et al., 2018, Dejonckere et al., 2001). Sound propagates in the air through variations in the air pressure typically measured in pascal (Pa), but also can be expressed as sound pressure levels in decibels (dB). SPL is the basic characteristic of voice and speech which has a logarithmic relationship with the sound pressure (p) defined by International Electrotechnical Commission-IEC 61672-1 (IEC, 2013) as:

$$\mathbf{SPL = 20\log_{10}\frac{p}{p_o}}\text{.........} \ \mathbf{(2)}$$

Where p_o is the reference for the sound pressure in air equal to 20 µPa (or 2×10^{-5}Pa). This reference value corresponds to the SPL of 0 dB, approximately the smallest sound pressure audible to human ear.

2.8.2.1 Measurement of voice SPL

The voice SPL can be measured using a sound level meter (SLM). It consists of an omnidirectional microphone, measurement circuitry and a display that shows SPLs with respect to the standard reference level. The three basic parameters set in the SLM include the level range, frequency weighting and time weighting or time averaging (Švec and Granqvist, 2018).

a) The *level range* sets the lowest and highest SPL limits that can be measured by keeping the total dynamic range constant, which is 20-100 dB, 30-110 dB, or 40-120 dB etc.
b) The *frequency weighting* approximates the sensitivity of human ear to softsounds at different frequencies. There are three types of frequency weightings, A, C and Z (**Figure 24**). A-weighting is the standard weighting of audible frequency response of human ear. The ear is not very sensitive to low frequencies but much sensitive to frequencies between 500 Hz to 8000 Hz. Similarly, the A-weighting covers the entire range of human hearing (20 Hz to 20 kHz) but shapes the filter to approximate the frequency sensitivity of human hearing. Therefore, the frequencies between 1-5 kHz are perceived loudest and strongest, while the low frequencies are attenuated. For example the 100 Hz sound is perceived 20 dB weaker than 1000 Hz. When A-weighting is used for measuring the sound pressure level, the results are depicted using dB(A) unit. The C-weighting is considered to be similar to the sensitivity of the human ear for loud sounds. The C-weighting response is relatively flat, within 3 dB, in the range of 32-8000 Hz. The frequencies below and above this range are attenuated (IEC, 2013). This type of weighting allows filtering out the ambient noise at frequencies outside the human voice frequency range and also eliminates the steady bias

voltage (DC component) present in the microphone (Švec and Granqvist, 2018). Measurements made using C-weighting are sometimes indicated by the unit dB(C).
The Z-weighting (zero or no weighting) has a flat response with the same sensitivity to all audible frequencies. Measurements made using Z-weighting are sometimes indicated by the unit dB(Z).

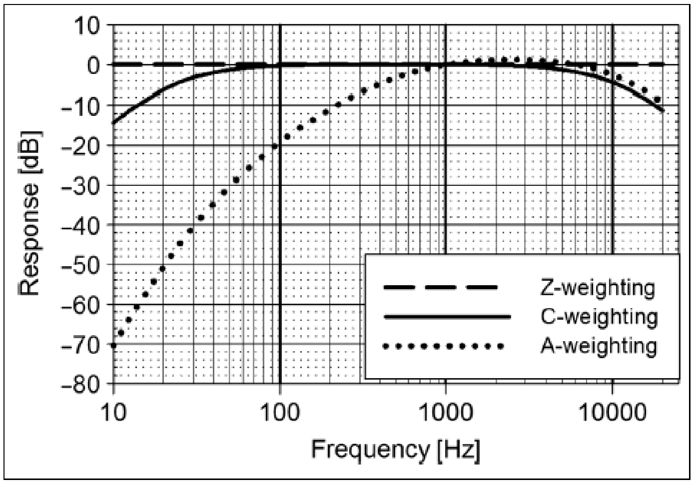


Figure 24: Standard frequency weighting curves for sound level measurement according to IEC 61672 (IEC, 2013). Taken from Švec and Granqvist (2018).

c) The *time weighting/averaging* is related to how rapidly the measurements respond to changes in the sound pressure. There are three types of time averaging measures: the time-weighted sound levels, time-averaged (equivalent) sound levels and peak sound levels based on instantaneous sound levels (Švec and Granqvist, 2018). In the present thesis, the *time-averaged (equivalent)* sound level of voice signal (or L_{eq}) is used in the **manuscript III** which is calculated using equation (2) where the p represents the root mean square (RMS) sound pressure over the specific time.

- In the present thesis, the teachers' voice SPL values reported in the **manuscript III** are the time-averaged (equivalent) C-weighted sound levels.

2.8.2.2 SPL Calibration

When we measure the recorded voice signal level, it is often affected by the computer software programs that produce relative but not calibrated SPL values of the voice signal. Thus calibration is required to enable measurements of absolute SPL values. There are three basic methods for SPL calibration (Švec and Granqvist, 2018). In the *first method* **(Figure 25a)** the calibration is done using a calibrator fitted on the SLM microphone, which produces the calibration tone of known SPL (example 94 dB). After that the calibrator is removed, and the calibrated SPL is announced to the microphone, which is also recorded in the file. Following this the voice/speech recording is done with a standard mouth to microphone distance, for example 30 cm as recommended by Švec and Granqvist (2010) which is also reported for further relating the measured SPL to vocal power. In the *second method* **(Figure 25b)** if the calibrator is not available, the calibration can be done using an SLM and loudspeaker. In this method, the microphone is positioned next to the SLM and both are at same distance from the loudspeaker. The recording is started and the calibration signal from the loudspeaker is generated and captured by the microphone, and the level read from the SLM is announced during the recording. After this the loudspeaker is removed and replaced by the subject and the voice recroding is carried out. The frequency weighting used in the SLM and the mouth to microphone distance during voice recording should be noted for its further use in calibration software. In the *third method* **(Figure 25c)** the SLM is positioned near the microphone at the same distance from the mouth. The recording is started and a person phonates a steady sustained vowel, and the level in the SLM is reported. The frequency and time weighting of SLM is also noted for its use in calibrating the measured levels in the software.

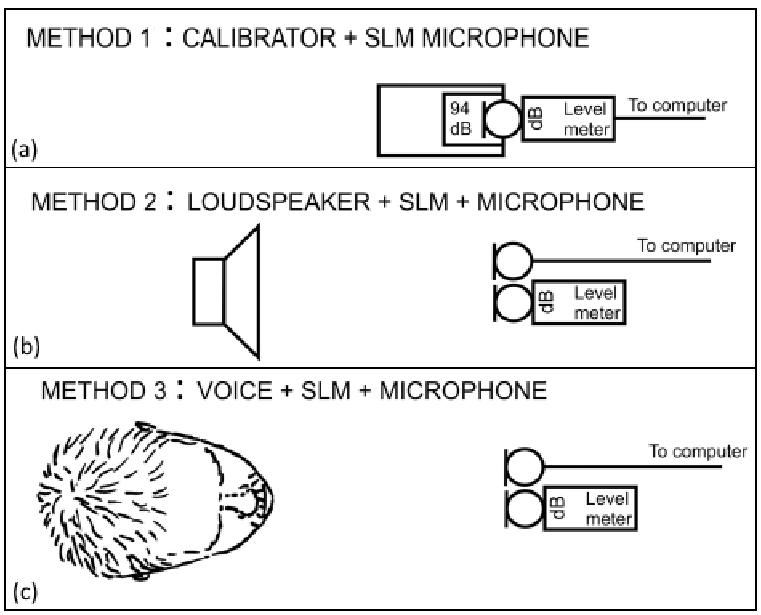


Figure 25: Three methods of SPL calibration for voice recordings. Figure modified from Švec and Granqvist (2018)

After having loaded the calibration tone and voice recording in a software, the L_{eq} of the calibration tone is to be checked. If the software indicated level is not same as that of the true SPL measured using the SLM, then there is a need for calibration of the levels in the software. In the present thesis, in **manucript III**, software *Praat* (Boersma and Weenink, 2013) was used for measuring the voice SPL. Here, for calibration, the measured levels are mathematically amplified to obtain the true sound pressure levels (that corresponded to the waveform values in pascals) using the multiplication factor $10^{\left(\frac{\Delta L}{20}\right)}$ where ΔL is the difference level (difference between the true sound pressure level read in the SLM and the uncalibrated level depicted in the *Praat* software). On completing this procedure, the calibration tone will show the true SPL values, and the voice/speech levels are said to be calibrated to the true SPLs.

- In the present thesis, measurement of voice SPL has been carried out in the **manuscript III**, along with cepstral measurements of teachers' voice. Previously there have been some indications of relationship between cepstral measures and vocal loudness in literature (Awan et al., 2012). Manuscript **III** explores this relationship in more detail.

- The next half of the thesis is experimental part, which summarizes the original work of the author (manuscripts **I, II, III** and **IV**) and provides an overall conclusion, with published papers attached at the end.

3 ORIGINAL WORK OF AUTHOR

3.1 Aims of the thesis

The present thesis is divided into two parts. The first part focuses on the videokymographic evaluation of vocal folds (**manuscript I** and **II**) and the second part investigates the teachers' voices (**manuscript III** and **IV**). Specific aims are:

1. **Manuscript I:** To evaluate the clinical value of VKG as an additional tool to stroboscopy for assessing and treating of various voice disorders and evaluate the diagnostic confidence of the clinician before and after VKG examination.
2. **Manuscript II:** To investigate the parameters that can be helpful to objectively quantify the 'lateral peak sharpness' parameter from the VKG images. The sharpness of the lateral peak is related to the vertical phase differences and mucosal waves on the vibrating vocal folds (**see introduction section 2.7.2.3.3**), providing useful information on the health and pliability of vocal fold mucosa.
3. **Manuscript III:** (a) To determine representative non-smoothed and smoothed cepstral peak prominence (CPP and CPPS) and voice SPL values in teachers, who consider themselves to have normal voice but some of them present with laryngeal pathology, (b) to investigate the changes of CPP, CPPS, SPL and perceptual ratings (vocal quality and firmness) for three voice tasks (comfortable vowel, comfortable speech and loud speech), and (c) study the influence of vocal pathology on the acoustic measures and perceptual ratings .
4. **Manuscript IV:** To investigate the influence of noise resulting from inapproprite location of schools, as well as the location and conditions of classrooms on vocal health of teachers.

3.1.1 Manuscript I: Clinical Value of Videokymography (VKG)

3.1.1.1 Introduction

Previously, videostroboscopy and HSV have been reported to be clinically useful in diagnosing some voice disorders (Woo et al., 1991, Paul et al., 2013, Sataloff et al., 1991, Patel et al., 2008, Braunschweig et al., 2008, Mehta et al., 2010b). However, due to the inherent limitations of these techniques (low sampling rate of videostroboscopy, and high time consumption of HSV along with its lack of immediate visual feedback and large amount of data requiring higher storage space) VKG was developed. The clinical value of VKG has not yet been formally evaluated, however, and therefore the present study was designed to evaluate the clinical usefulness of VKG as an additional method to stroboscopy for assessing voice disorders. Additionally, the clinician's confidence in the diagnosis (after stroboscopy and after VKG examination) was also evaluated. This study was carried out in collaboration with the clinicians from Medical Healthcom, Voice Center Prague, Czech Republic.

3.1.1.2 Questions addressed

1. Does VKG help in establishing the diagnosis of voice disorders in patients?
2. Does VKG improve the clinician's confidence in diagnosis previously based on stroboscopy?
3. Does VKG help in treatment decisions?

3.1.1.3 Materials and Methods

1. *Participants:* The participants for the study were 105 individuals with voice complaints. (These were 71 females and 34 males, aged between 10-80 years. Out of 105 individuals, 42 participants were professional voice users. The participants primarily complained of hoarseness, vocal fatigue, and loss of vocal range. All the participants underwent routine ENT procedures, including case history, ear, nose and oral cavity examination.

2. *Questionnaire and procedure:* For evaluating the clinical value of VKG, a questionnaire (**see supplementary material of manuscript I**) was designed with two parts. The first part focused on the diagnostic contribution of stroboscopic examination, and the second part focused on the diagnostic contribution of VKG as an addition to stroboscopy. A senior laryngologist with 30 years of experience in ear-nose-throat (ENT) practice and more than 10 years of experience in working with VKG, performed the two examinations on all the patients and filled in the questionnaire.

   The following steps were carried out:

   Step 1: After the routine ENT evaluation, the clinician performed stroboscopic examination on the patient.

   Step 2: After stroboscopic examination, the clinician filled in the first part of the questionnaire which specified the diagnosis, treatment recommendations and confidence in the diagnosis based on stroboscopy. The diagnosis was made based on the International Classification of Diseases (ICD 10 Version: 2016) (**recall Table 1 of introduction section 2.2**), and the clinician rated her diagnostic confidence based on a five-point rating scale: 0-not confident; 1-little confident; 2-moderately confident; 3-greatly confident; 4-absolutely confident.

   Step 3: The clinician performed the VKG examination on the same patient and filled in the second part of the questionnaire thus providing a final diagnosis and treatment recommendations based on VKG.

   Step 4: The clinician listed the VKG features that were useful in reaching the final diagnosis, and rated the usefulness of VKG on a five-point rating scale: 0-no diagnostic contribution of VKG; 1-VKG confirmed the stroboscopic diagnosis; 3-VKG made the diagnosis more accurate and resulted in an adjustment of offered treatment; and 4-VKG changed the initial diagnosis and changed the offered treatment.

Step 5: The clinician also rated her diagnostic confidence after VKG examination on the same rating scale as used for stroboscopy.

3.1.1.4 Results

1. *VKG usefulness:* The results of the ratings of the clinical value of VKG are shown in **Figure 26**. In 95 % of cases, VKG was found to be useful for establishing the diagnosis, while in 5 % of individuals VKG was rated not useful. In no case (0 %) the VKG evaluation was found to be critical by completely changing the stroboscopic diagnosis and treatment recommendations. **Table 10** shows the distribution of the diagnosis of study participants and the ratings (five-point rating R0 to R4) for the diagnostic usefulness of VKG.

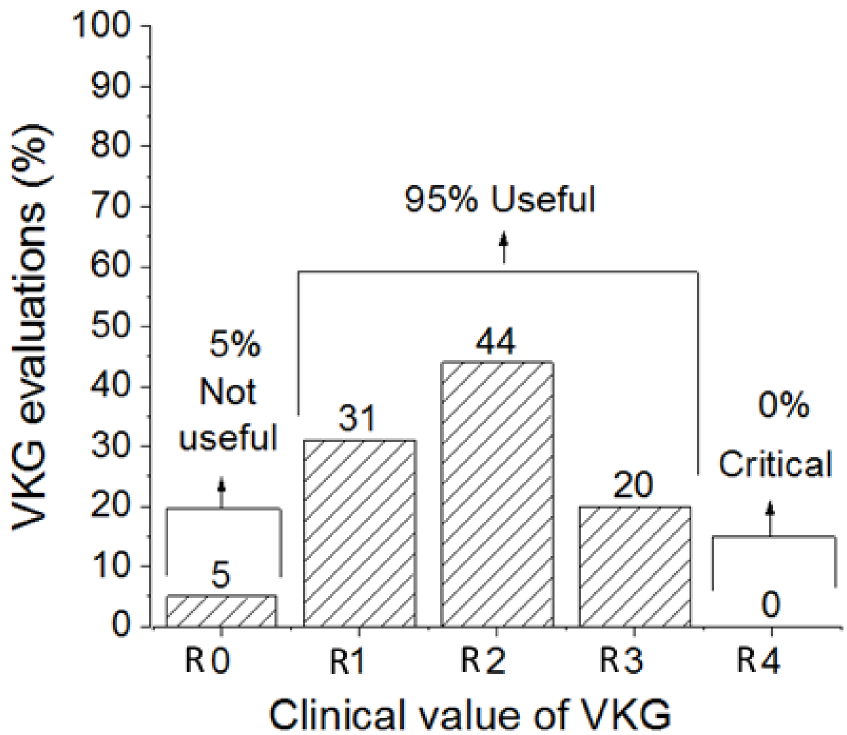


Figure 26: The clinical value of VKG in 105 evaluations. 5 point rating scale (R0-no diagnostic contribution of VKG; R1-VKG confirmed the stroboscopic diagnosis; R2-VKG made the diagnosis more accurate; R3-VKG made the diagnosis more accurate and resulted in an adjustment of offered treatment; and R4-VKG changed the initial diagnosis and changed the offered treatment).

Table 10: Distribution of diagnoses of study participants and of the ratings for diagnostic usefulness of VKG.

Diagnosis	Cases	Distribution of ratings of usefulness (cases)				
		R0	R1	R2	R3	R4
Laryngitis (chronic & acute, incl. reflux)	65	3	16	29	17	-
Hyperfunctional dysphonia	10	1	5	4	-	-
Vocal fatigue	5	-	2	1	2	-
Vocal fold paralysis	4	-	2	2	-	-
Vocal polyp	4	-	-	4	-	-
Larynx and vocal fold edema	4	-	1	3	-	-
Singing technique problems with normal laryngeal findings	3	-	3	-	-	-
Nasopharyngitis with normal laryngeal findings	2	-	1	-	1	-
Vocal nodules	2	-	-	1	1	-
Mutational voice disorder	2	-	1	1	-	-
Dilated blood vessels	1	-	-	1	-	-
Vocal fold carcinoma	1	-	1	-	-	-
Spasmodic dysphonia	1	1	-	-	-	-
Asthma bronchiole and coughing exacerbation	1	-	1	-	-	-
Total (No.)	105	5	33	46	21	0
Total (%)	100	5	31	44	20	0

2. *Useful VKG features:* **Table 11** lists the most important VKG features that helped the clinician to reach the final diagnosis. The features are listed in the order of most frequently occurring to least frequently occurring features.

Table 11: The occurrence of VKG features that were listed as helpful for the diagnosis in the 105 cases.

VKG Features	Cases
Rounded lateral peaks	58 %
Missing or reduced mucosal wave	44 %
Left-right phase differences	27 %
Reduced vocal fold amplitude	26 %
Missing glottal closure	16 %
Normal VKG findings	15 %
Missing vocal fold vibration	8 %
Irregular vocal fold vibration	5 %
Glottal closure too short	3 %
Sharp medial peaks	3 %
Left-right frequency differences	2 %
Interference of vibrating ventricular folds and surrounding structures	2 %

3. *Clinician's diagnostic confidence:* After the VKG examination, the diagnostic confidence of the clinician improved in 68 % cases and remained the same in 32 % cases.

3.1.1.5 Discussion

VKG was found to be useful in 95 % of cases (including individuals with laryngitis (chronic & acute, incl. reflux), hyperfunctional dysphonia, organic lesions, vocal fatigue, and those with upper respiratory infections) (**Figure 26, Table10**). Out of these 95 % cases, VKG confirmed the initial stroboscopic diagnosis in 31 % cases, not providing any further insights into the nature and degree of the disorder. In 44 % cases, VKG made the stroboscopic diagnosis more accurate by assessing the extent and severity of the vocal fold pathology. For example in cases of laryngitis, VKG was useful in assessing how much the pathology had affected the pliability of vocal fold mucosa. In about 20 % cases the VKG altered the treatment offered, in terms of refinements in the voice therapy (4 % cases), change in medication (3 %), modification in dosage of drugs (10 %), and clarification in duration of voice rest (3 %). In the remaining 5 % cases, VKG was not useful, as it did not indicate any other pathology that would otherwise be missed on stroboscopy. Also in no cases (0 %) VKG was found critical by completely changing the stroboscopic diagnosis and treatment decisions. VKG was primarily found useful for bringing more insights into the seriousness of the voice disorder based on the abnormal vibratory behavior of the vocal folds rather than providing a structural diagnosis such as polyp, nodule, or cyst etc.

Considering the useful VKG features listed in **Table 11**, the most frequently occurring VKG features were rounded lateral peaks and missing or reduced mucosal waves seen in 58 % and 44 % of subjects respectively. Both these features are related to reduced mucosal pliability and increased stiffness of the vocal fold mucosa resulting from vocal fold pathology (Švec et al., 2009, Yamauchi et al., 2016a). Other VKG features that were reported to be useful were left-right phase difference (27 %) and amplitude difference (26 %). These left right asymmetries are usually a result of changes in the mass, tension and stiffness of the vocal folds (Isshiki et al., 1977, Zhang and Hieu Luu, 2012, Pickup and Thomson, 2009, Švec et al., 2007b). Such asymmetries may be seen in individuals with voice disorders (Yamauchi et al., 2015b, Isshiki et al., 1977, Bonilha et al., 2012) as well as in normal cases (Bonilha et al., 2008). The remaining VKG features listed in **Table 11** were less frequently occurring in our subjects and therefore require investigating its usefulness on a large variety of voice disorders.

After the VKG examination, the clinician's diagnostic confidence improved in 68 % cases. The clinician had additional information pertaining to the abnormal vibratory features especially in individuals who had complaints of hoarseness, but presented with normal stroboscopic findings. VKG results helped the clinician to perform further diagnostic investigations in order to find the cause of the abnormal vocal fold vibrations.

3.1.1.6 Conclusion

Based on the results of this study, it was concluded that VKG was found to be useful for taking some important diagnostic and treatment decisions especially when patients present with indecisive laryngoscopic and stroboscopic findings, where clinicians are uncertain about their diagnosis and treatment.

3.1.2 Manuscript II: Quantifying 'lateral peak sharpness' from VKG images

3.1.2.1 Introduction

In the previous study (**manuscript I**), the most frequently occurring and helpful VKG feature for diagnosis was the rounded lateral peak shape that was evaluated based on visual observation of VKG images. This feature gave important diagnostic information pertaining to the health and pliability of vocal fold mucosa. In the present study (**manuscript II**), objective parameters that can be helpful in quantifying the lateral peak sharpness from the VKG images are investigated. Quantification of this parameter will help clinicians to have a reliable standard reference value, based on which better diagnosis and treatment decisions can be taken.

3.1.2.2 Objectives

1. To visually evaluate pre-selected clinical VKG images to obtain subjective ratings of the lateral peak sharpness.
2. To obtain contours of the vibrating vocal folds as waveforms from the above selected images through automatic image analysis procedure.
3. To quantify the waveforms in order to obtain parameters expected to reflect the lateral peak sharpness.
4. To correlate the quantified parameters with visual ratings, in order to determine the best parameter that represents sharpness of lateral peaks.

3.1.2.3 Materials and methods

1. *Dataset:* 45 VKG images were retrospectively selected from clinical records of patients examined for voice complaints at the Voice and Hearing Centre, Medical Healthcom, Ltd., Prague, Czech Republic. These images demonstrated varied degrees of sharpness of lateral peaks.

2. *Visual rating:* Three raters evaluated the shape of the lateral peaks for both right and left vocal folds separately from the VKG images based on a 4 point rating scale (1-sharp; 2-rather sharp; 3-rather rounded; 4-rounded) by using a pictogram representation of different sharpness of lateral peak as described in introduction section, **Figure 14 e** (Švec et al., 2007c). For assessing the intra-rater reliability, each rater performed the evaluation twice

with a pause of 7-10 days in between. During the second evaluation, the order of the images was changed to minimize the memory effect. The ratings from the two evaluations for the three raters were consolidated and an average (visual average) was obtained. A common consensus (visual consensus) was also arrived through discussion among the three raters due to presence of some discrepancies between raters.

3. *Image Analysis:* The VKG analyzer software (**see introduction section 2.7.2.3.4 under VKG**) was used to extract the glottal edge contours of the vocal folds, saved in text files. A custom MATLAB script (MATLAB R2016a) was used to process the vocal fold contours and to obtain parameters revealing on the lateral peak sharpness.

4. *Quantification of Lateral Peak Sharpness:* Two sets of parameters *(1) the Open Time Percentage Quotients* (OTQ) and (2) *Plateau Quotients* (PQ) were defined and quantified to check their capability of reflecting the sharpness of the lateral peak.
   a) Open time percentage quotient (OTQ_R) is the proportion of time during which the vocal fold displacement exceeds R % of vibration amplitude within a period (**Figure 27**). It is defined as:

$$\mathbf{OTQ_R} = \frac{\mathbf{D_R}}{\mathbf{T}}$$

   where D_R is the duration of phase when lateral displacement is greater than R % of the vibration amplitude and T is the period of the vocal fold vibration cycle.

   b) Plateau quotient (PQ_R) is the proportion of time during which the vocal folds displacements exceed R % of vibration amplitude within an open phase (**Figure 27**). It is defined as:

$$\mathbf{PQ_R} = \frac{\mathbf{D_R}}{\mathbf{OP}}$$

   where OP is the duration of the open phase.

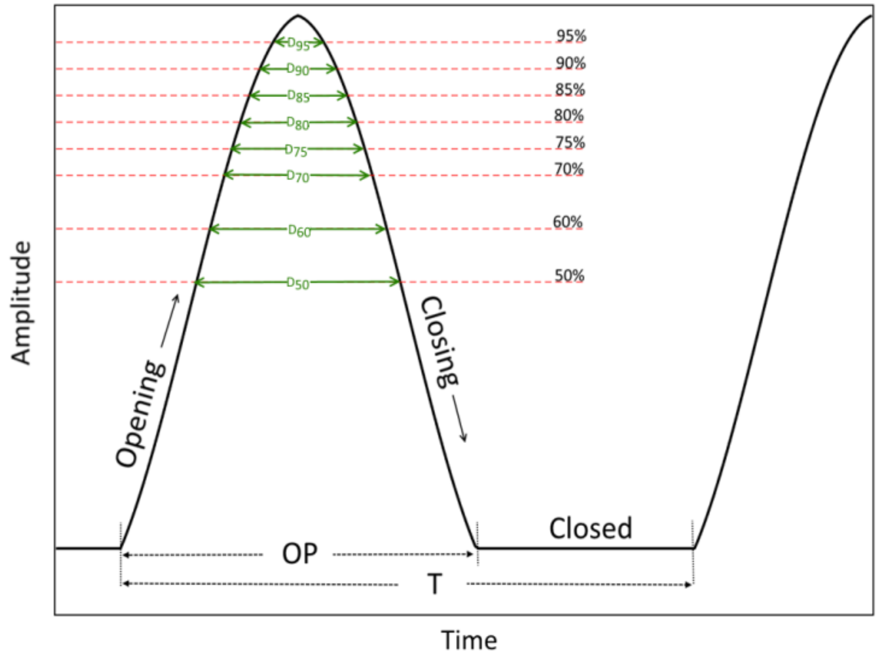


Figure 27. Parameterization of the vocal fold waveform to obtain the Open Time Percentage Quotients (OTQ_R) and Plateau Quotients (PQ_R) revealing on the peak sharpness. OP is the open phase, T is the period and D_R are the durations of the phases during which the waveform is above specified percentage *R* of the amplitude. The R percentages are indicated by the dashed lines.

3.1.2.4 Results

1. *Reliability of visual ratings:* The Spearman's rank correlation coefficients for the individual raters varied from 0.84 to 0.85 (P<0.001, N=90) indicating very strong and significant correlations between the repeated evaluations. The inter-rater reliability measured by Spearman's r varied from 0.67 to 0.82, with the mean value of r=0.73; N=90. These values were also found to be strong and significant (P<0.001). Also, strong correlation was obtained between the visual consensus and visual average (Spearman's rank correlation r =0.99, P<0.001).

2. *Correlations between visual ratings and the analyzed parameters:* The correlations between different OTQ and PQ parameters with the visual consensus and visual average ratings are shown in **Figure 28.** All the OTQ and PQ parameters measured at different R % had significant correlations (P<0.001) with both the visual ratings (average and consensus). Among these, the highest correlations (indicated by red arrows in **Figure 28**) were obtained for the parameters measured at 95 % amplitude (OTQ_{95}, PQ_{95}) and at 80 % amplitude (OTQ_{80}, PQ_{80}). Lowest correlations were found for the parameters measured at 50 % the amplitude (OTQ_{50}, PQ_{50}).

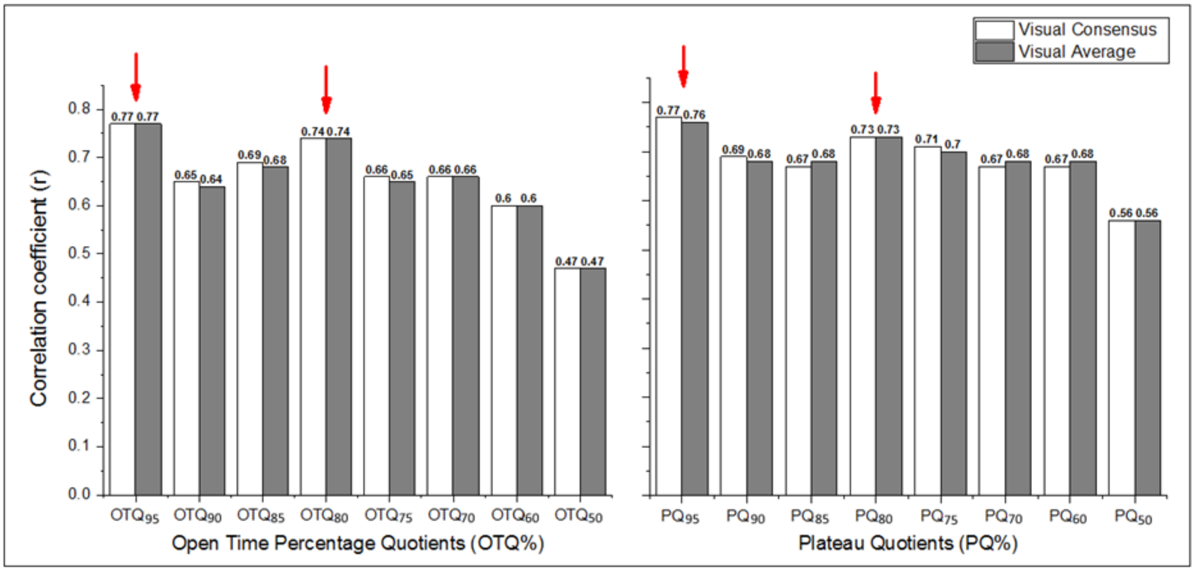


Figure 28. Spearman's rank correlation coefficients indicating the agreement between the visual ratings and the measured parameters OTQ and PQ. The highest correlation coefficients are indicated by red arrows.

The regression plots between the selected OTQ ($OTQ_{95,}$ OTQ_{80}) and PQ (PQ_{95}, PQ_{80}) parameters and visual consensus rating are shown in **Figure 29**. It is clear from the graph that all the quotients increase in their values when the shape of the lateral peak changes from sharp to rounded. However, some discrepancies exist between the visual and automatic evaluations as seen by some spread of the measured data around the best fit line.

3.1.2.5 Discussion

Considering the results of the reliability of visual ratings, the inter-rater reliability was slightly lower than the intra-rater reliability indicating more disagreements between the visual evaluations of different raters than between repeated evaluations of the same rater. When these ratings were correlated to the quantified OTQ and PQ parameters at different R % values, strong correlation was obtained for the OTQ and PQ parameters derived at 95 % and 80 % amplitude and worst correlations appearing at 50 % amplitude (**Figure 28**). Since the peak corresponds to 100 % of amplitude, the values obtained closer to the peak should theoretically show better correlations than those far off from the peak. However, the correlations at 90 % and 85 % amplitude were worse than those at 80 % because of the presence of contour artifacts due to the limited pixel and time resolution (**see Figure 6 in manuscript II for details).** Also the discrepancies or outliers present in the **Figure 29** of regression plots may also be due to contour detection artifacts occurring due to image analysis procedure.

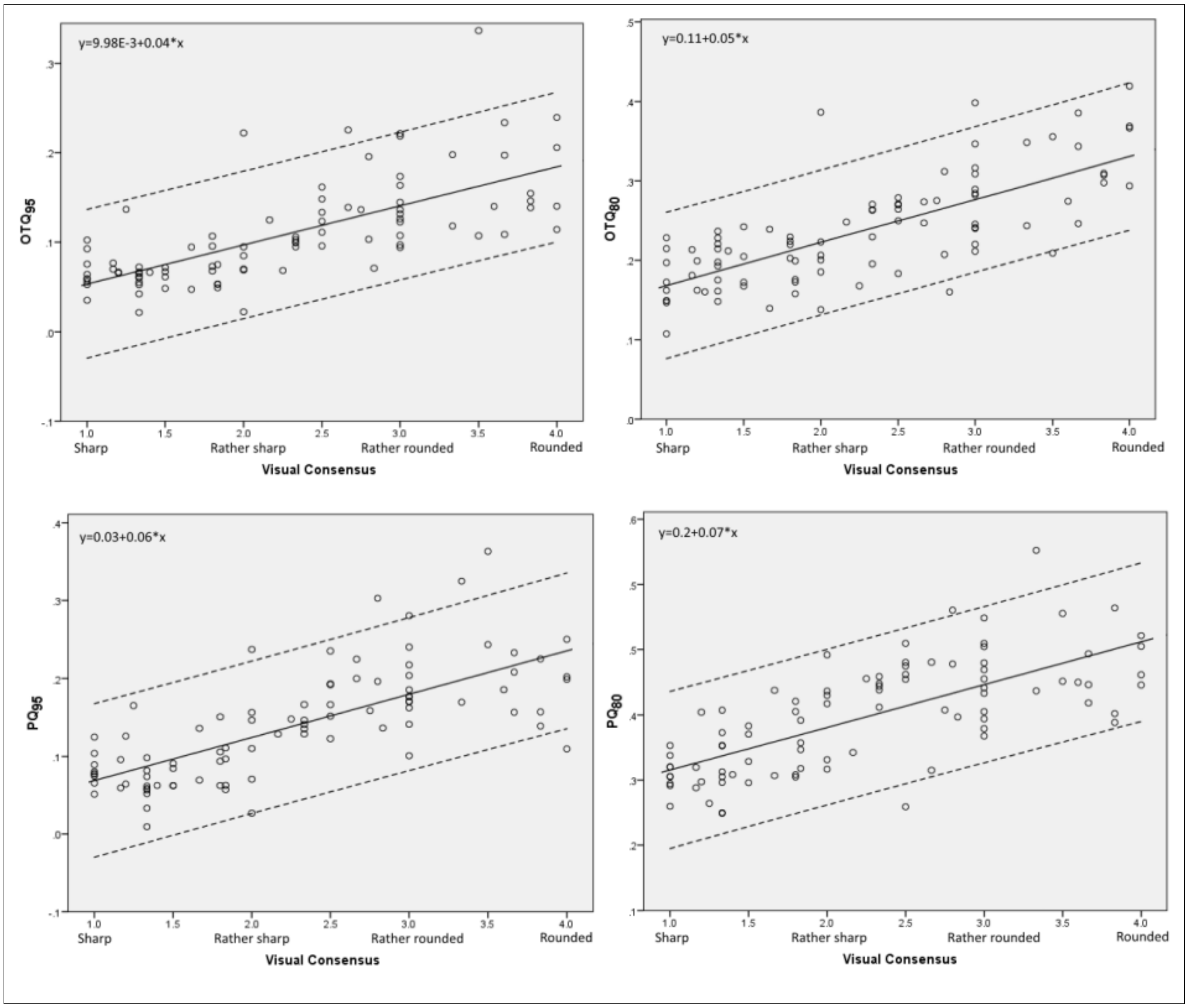


Figure 29. The relationship between the measured values and the visual ratings for the four parameters with the highest correlations – OTQ_{95}, OTQ_{80}, PQ_{95} and PQ_{80}. The lines indicate the best fit linear relationship (solid) and 95 % confidence intervals (dashed).

3.1.2.6 Conclusion

The PQ_{95}, PQ_{80}, OTQ_{95} and OTQ_{80} parameters stood out as the possible candidates for capturing the sharpness of the lateral peaks. The reliability of these parameters appears comparable to the inter-individual reliability of visual ratings. The results provide basic insights into developing the computer algorithms to automatically quantify the sharpness of lateral peaks from the VKG images.

3.1.3 Manuscript III: Cepstral and perceptual analysis of teachers' voice

3.1.3.1 Introduction

Cepstral analysis of voice has been considered to be a measure of overall severity of dysphonia (Heman-Ackah et al., 2003, Maryn et al., 2009), and also been reported to correlate well with perceptual evaluation of different dysphonia types (hoarse, breathy, pressed, and strained voice types) (Wolfe and Martin, 1997, Awan et al., 2010). The present study investigates the influence of three different phonatory tasks (comfortable vowel, comfortable speech and loud speech) on CPP, CPPS and voice SPL values and perceptual ratings of vocal quality and firmness of voice in teachers who consider themselves to have normal voice, but some present with laryngeal pathologies detected laryngoscopically. The impact of these underlying vocal pathologies on the acoustic and perceptual measures has also been studied.
Since teachers often indulge in speaking at increased loudness levels in presence of unfavorable environmental conditions leading to vocal fatigue, strain and increased voice symptoms, the present study also included analysis of loud speech samples, to assess its influence on the acoustic (cepstral and voice SPL) and perceptual measures (voice quality and firmness rating).

3.1.3.2 Materials and Methods

1. *Participants and laryngeal status:* A total of 84 Finnish female primary school teachers volunteered as subjects for this study. All the participants considered themselves to be vocally healthy and capable of carrying their profession, however some laryngeal changes were found in 44 (52.4 %) teachers, where 33 of them (39.3 %) had mild changes (mild vocal fold erythema, arytenoid erythema, mild edema and mild glottal closure insufficiency) and 11 (13.1 %) had substantial changes (nodules, polyps, chronic laryngitis, laryngeal reflux disease, and moderate to severe glottal closure insufficiency). Laryngeal status evaluation was done by an experienced phoniatrician on a three point scale (1-healthy; 2-mild changes; 3-disorderd), after a case history and indirect mirror laryngoscopy.

2. *Tasks:* Teachers were asked to sustain a prolonged vowel [a:] for 5 seconds, followed by reading of a text (two sentences from a text passage from Saroyan W: The human comedy. Harcourt 1943. In Finnish: Ihmisiä elämän näyttämöllä. Tammi 1959) at comfortable loudness as in conversational speaking. Additionally, the teachers were asked to read the same text at an increased loudness level as if teaching in a large noisy classroom.

3. *Recordings and acoustic measurements:* The voice recordings of the tasks were carried out in primary schools, in teacher's own classrooms with minimal ambient noise (approximately about 35 dB(A)). Recordings were made using a portable digital recorder and an omnidirectional head-mounted microphone selected according to the recommendations by Švec and Granqvist (2010). The microphone was maintained at a constant distance of 6 cm,

at an angle of 45^{o} from the side of the subject's mouth. The voice recordings were then calibrated to obtain the true sound pressure level (SPL) of vowel and speech samples in *Praat* software based on the guidelines by Švec and Granqvist (2018). The final single SPL value obtained was a close approximation of the time-averaged (equivalent) C-weighted sound level for the entire voice sample selected as measured by the sound level meter. The CPP and CPPS values for all the three tasks were extracted using software *Praat* (version 5.4.05) with *Praat* default settings and settings recommended by Maryn and Weenink (2015) (**see manuscript III, Table 2 and introduction section 2.8.1.4, Table 8**).

4. *Perceptual analysis:* The same samples of comfortable vowel phonation and comfortable and loud speech reading that were analyzed for cepstral measures were also perceptually analyzed by five experienced voice experts. They rated overall voice quality along a ten point unipolar scale from 0 = poor to excellent = 10. Additionally, they evaluated the vocal firmness along a bipolar axis from 0 = breathy through 5 = adequate to 10 = pressed. The individual listeners' ratings were averaged for each sample to be used in the statistical analyses.

3.1.3.3 Results

1. *Acoustic and perceptual results for three voice tasks:* **Table 12** shows that the CPP, CPPS and SPL values were significantly larger for both comfortable sustained vowels and loud speech than for speech at comfortable loudness ($P<0.001$). Perceptually, vowels had better voice quality than comfortable speech. Loud speech was perceived to be firmer and have a better voice quality than comfortable speech ($P<0.001$).

Table 12: The evaluation results expressed through the mean and standard deviation values for the three voice samples.

Voice samples	CPP (dB)	CPPS (dB)	Voice SPL (dB)	Voice quality	Vocal Firmness
Sustained vowel	23.4±2.9	13.6±2.1	82.4±5.5	4.7±0.9	5.1±1.0
Comfortable speech	19.0±1.4	10.4±1.5	76.4±3.3	4.4±1.0	4.8±1.2
Loud speech	19.6±1.2	11.4±1.4	84.9±3.8	4.9±1.0	5.7±1.2

2. *CPP and CPPS versus SPL for vowel and speech:* Significant correlations were obtained between SPL and cepstral measures (both CPP and CPPS) for vowel samples. No significant correlation was obtained between SPL and cepstral measures for comfortable speech. However, significant correlations were obtained between the acoustic measures when both comfortable and loud speech data were pooled together. **Figure 30** shows the relationship between the voice SPL and cepstral measures for the vowel and pooled speech data in more detail. For vowel, with every 10 dB increase in SPL, CPP increased by 2.4 dB and CPPS by

1.8 dB. Also for the pooled speech sample, for every 10 dB increase in SPL, CPP increased by 0.7 dB and CPPS by 1.2 dB.

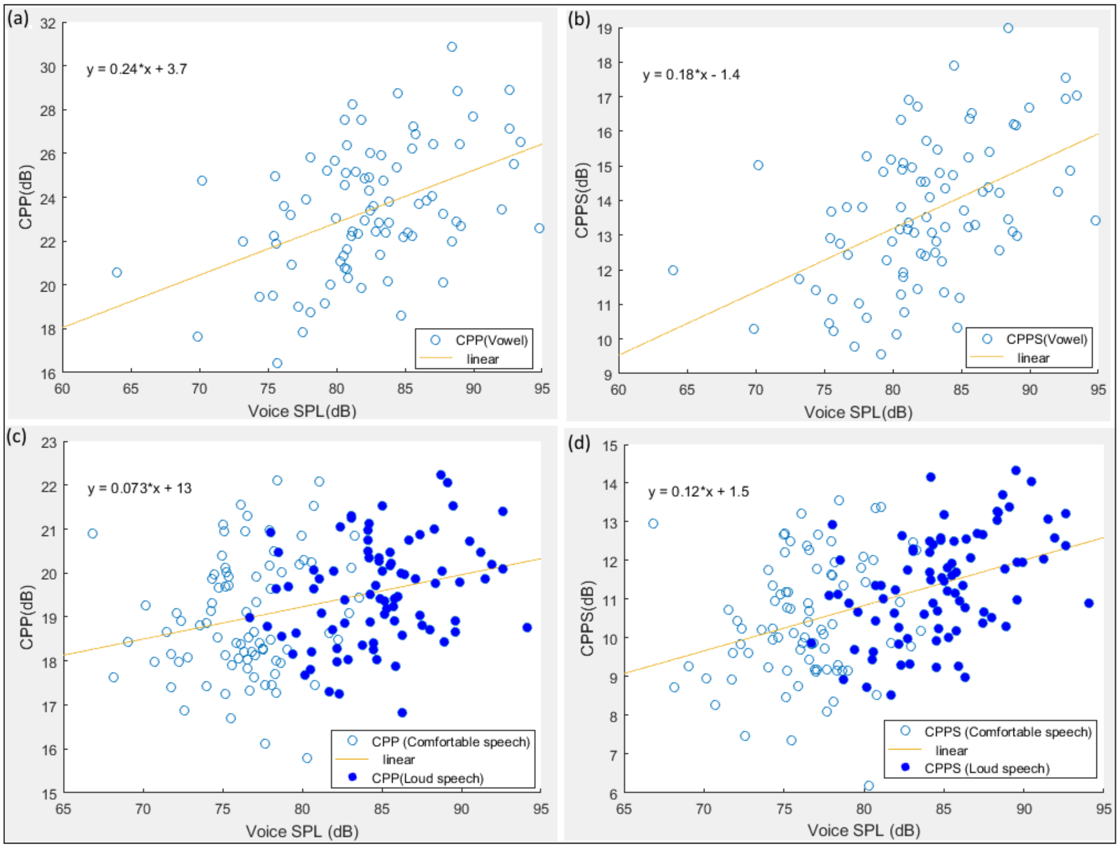


Figure 30: A scatterplot showing relationship between time-averaged equivalent SPL (at 6 cm distance in dB re 20 μPa) and the two cepstral measures for sustained vowel [a, b] and speech [c, d]. The speech data contain both the comfortable (empty circles) and loud (filled circles) conditions together. Notice the linear regression lines with their equation shown in each of the graphs – all of them show the trend of CPP/CPPS increase with increased SPL.

3. *Voice perception versus laryngeal pathology:* No significant correlations were found betwen perceptual ratings and laryngeal status findings for any of the vocal tasks. Also ANOVA showed no significant differences across the three laryngeal status categories (P>0.05) for the perceptual ratings of voice quality and firmness.

4. *Acoustic measures (CPP, CPPS and SPL) versus laryngeal pathology:* No significant correlations were found between acoustic measures (CPP, CPPS or SPL) and laryngeal status categories, for the three tasks. Also no significant differences for the acoustic measures across the three laryngeal status category was seen (P>0.05). Nevertheless, there was systematic decrease of the CPP, CPPS and SPL values from healthy to mild to disordered category in all the three voice samples. The numerical results are shown in **Table 13**.

Table 13: Mean, standard deviation (SD) and standard error (SE) of cepstral values (in dB) and of the time-averaged equivalent SPL at 6 cm distance (in dB re 20 μPa) for all three voice samples for teachers grouped under the laryngeal status category.

laryngeal status category	N	Vowel			Comfortable speech			Loud speech		
		CPP	CPPS	SPL	CPP	CPPS	SPL	CPP	CPPS	SPL
		Mean ±SD (SE)	Mean ±SD (SE)	Mean ±SD (SE)	Mean ±SD (SE)	Mean ±SD (SE)	Mean ±SD (SE)	Mean ±SD (SE)	Mean ±SD (SE)	Mean ±SD (SE)
Healthy	40	23.8 ±2.6 (0.41)	13.9 ±1.9 (0.30)	83.2 ±6.1 (0.97)	19.0 ±1.5 (0.23)	10.5 ±1.2 (0.25)	77.0 ±3.0 (0.47)	19.7 ±1.3 (0.20)	11.5 ±1.5 (0.23)	85.0 ±3.8 (0.60)
Mild	33	23.0 ±3.4 (0.59)	13.4 ±2.4 (0.41)	81.8 ±5.2 (0.90)	19.0 ±1.2 (0.21)	10.4 ±1.5 (0.26)	76.2 ±3.9 (0.68)	19.5 ±1.1 (0.18)	11.3 ±1.2 (0.21)	85.0 ±4.1 (0.70)
Disordered	11	23.0 ±2.3 (0.68)	13.3 ±1.6 (0.49)	81.0± 3.5 (1.06)	18.8 ±1.4 (0.43)	10.1 ±1.5 (0.44)	74.9 ±2.1 (0.62)	19.4 ±1.4 (0.32)	11.2 ±1.2 (0.38)	84.0 ±2.9 (0.88)

3.1.3.4 Discussion

Considering sections 1 and 2 of the results, CPP and CPPS values were significantly larger for comfortable sustained vowels than comfortable speech. This is likely due to the fact that speech samples consist of voice onsets and offsets, vocal pauses and variations in fundamental frequency and intensity etc., all of which decrease the prominence of harmonic organization over the noise content measured by the CPP and CPPS parameters (Zhang and Jiang, 2008, Maryn and Roy, 2012). Furthermore, due to presence of voiceless consonants, and pauses between words and sentences, there is an overall decrease in the voice SPL of speech compared to vowel samples. In case of loud speech, the regression plot **(Figure 30 (c &d)), data indicated by filled circles)** shows that with increase in the voice SPL values, the cepstral measures also increase. This linear relationship between voice SPL and cepstral measures may be due to decrease in the perturbation of voice, when phonational loudness increases (Brockmann et al., 2018), rendering an increase in the cepstral values (Awan et al., 2012). Also there is increase in medial compression of vocal folds and improvement in the glotal closure that decreases the glottal noise and increases the strength of overtones in the signal (Awan et al., 2012, Sulter and Albers, 1996).

Significant linear correlations were obtained between cepstral measures and voice SPL for vowel samples (**Figure 30**). No significant correlation between cepstral measures and voice SPL was seen for comfortable speech sample, because of smaller range of SPL for this voice task in comparison to rather large spread of CPP and CPPS values for different individuals (**recall Figure 30 (c & d), data indicated by empty circles only**). However, when the comfortable speech and loud speech data were pooled together, the SPL range enlarged rendering significant correlation between SPL and cepstral measures.

Considering perceptual evaluation, vowels were perceived to have better voice quality than comfortable speech, which is likely due to the fact that speech samples are more demanding on laryngeal movement coordination and expose voice abnormalities more extensively than sustained vowels (Law et al., 2012). Loud speech was perceived to be firmer and have a better voice quality than comfortable speech, which may be due to the increase in medial compression of vocal folds and improvement in glottal closure due to increase in voice loudness (Awan et al., 2012, Sulter and Albers, 1996). Also the mean values for firmness rating of loud speech was 5.7±1.2 (**Table 12**) indicating that the voices of the teachers were neither breathy nor pressed, and that they did not endanger their larynges with inappropriate vocal mechanism.

Our CPPS results (**Table 12**) are similar to those obtained by Maryn and Weenink (2015) and Latoszek et al. (2018) using *Praat*, but different from the CPPS results reported by Watts et al. (2017) and Sauder et al. (2017), also using Praat. However, the CPPS results reported by Watts et al. (2017) and Sauder et al. (2017) are similar to our CPP (unsmoothed) values. Therefore we suspect that the cepstral values reported by Watts et al. (2017) and Sauder et al. (2017) are not the smoothed CPPS values but rather the non-smoothed CPP ones. Differences in the values of CPP and CPPS are likely due to differences in the time and quefrency averaging window settings. This topic has already been elaborated in the introduction section **2.8.1.4 (Table 9**). Clinicians are required to check or adjust the parameter settings prior to performing cepstral analysis, inorder to have reproducible results.

Considering the result sections 3 and 4, neither the perceptual ratings, nor the acoustic measures significantly correlated or differed across the laryngeal status categories (healthy, mild and disordered). The voice raters could not perceive any vocal changes as a result of underlying pathology, similar to the teachers who did not complain of any voice problems. The difference in the values of acoustic measures and perceptual ratings between the healthy and disordered group was very small and not significant (**Table 13**). This indicates that the vocal pathologies did not have any significant influence on the perceptual and acoustic measures.

Nevertheless, the results in **Table 13** revealed that the acoustic measures for all three voice samples, showed a consistent decline in the mean values with increase in the severity of laryngeal pathology. This indicates that the acoustic measures (voice SPL, CPP and CPPS) may be more sensitive to the underlying vocal pathology than perceptual measures. But, the lowering of these values in disordered group needs more exploration in future studies because in this study, the pathologic group size was limited to only 11 teachers causing the standard error of the mean to be rather large for finding significant differences. Since the teachers did not have any voice complaints, the cepstral and voice SPL data from this study could be considered representative for teachers with *functionally healthy voice*. Furthermore, this data may be useful for comparisons with the values obtained for teachers who present with voice complaints and diagnosed with voice disorders.

3.1.3.5 Conclusion

Based on the results of this study, neither the acoustic measures (CPP, CPPS and SPL) nor the perceptual evaluations could clearly distinguish teachers with laryngeal changes from laryngeally healthy teachers. The acoutic data could be considered representative for teachers with functionally healthy voice, and future studies should check for the decreasing trend of CPP, CPPS and SPL with hidden pathology, on a larger study population of teachers with voice disorders.

3.1.4 Manuscript IV: Influence of noise on teachers' voice

3.1.4.1 Introduction

The influence of classroom acoustics on the vocal fold of teachers has been documented objectively in previous studies (Kristiansen et al., 2014, Puglisi et al., 2017, Durup et al., 2015). However, noise resulting from inappropriate location of schools and classrooms and poor classroom conditions and its negative impact on vocal health of teachers are often not investigated or not considered as an important causative factor for poor vocal health of teachers. Therefore this study documents the teachers' self-reports on severity and frequency of their voice symptoms and investigate its correlation with their reports on noise perception and the location and conditions of schools and classrooms.

3.1.4.2 Materials and methods

1. *Participants:* 140 teachers (85 females and 55 males) aged 21–56 years (mean age = 35.8 years) from primary and preparatory schools in Upper Egypt participated in this study. Out of the 69 teachers from primary schools, 36 taught in public schools and 33 in private. 71 teachers worked in preparatory schools, 34 of which taught in public schools and 37 in private. Public and private school teachers had an average of 17.9 years and 7.4 years of teaching experience respectively, with a total average of 12.3 years combining both the groups (Abo-Hasseba et al., 2017).

2. *Questionnaire:* Teachers were asked to fill out a questionnaire (in Arabic language) regarding their demographic data (age, gender, type of school taught at, total years of teaching experience), and regarding the frequency (1-no recurrence; 2-monthly recurrence; 3-weekly recurrence; 4-daily recurrence), and severity (1-none; 2-mild; 3-moderate; 4-severe) of various voice symptoms (dysphonia, laryngeal pain, throat clearing, throat dryness, voice interrupted by the end of the day, and extra voice effort required to continue speaking). They were also asked about their school and classroom location and conditions. The teachers also reported their perception on existing noise (and also the source of noise) at their work place, and how they felt about it (feeling of being in a noisy environment and having to raise their voice due to noise; both rated on a four-point rating scale (1-always; 2-sometimes; 3-rarely; 4-never).

3.1.4.3 Results

1. *Relationship between teachers' self-reported noise source and voice symptoms:* There were significant correlations between the teachers' self-reported noise sources and voice symptoms on chi-squared test (P<0.001). Teachers experienced frequent laryngeal and neck pain, that was daily to monthly recurring, and they also had to increase their vocal effort in order to talk for longer durations. These symptoms were mostly related to presence of noise from student activities and talking in their own classrooms (reported by 61.4 % teachers) and noise from neighboring classrooms (52.9 % of teachers).

2. *Relationship between teachers' noise perception and voice symptoms:* About 57.9 % teachers felt that they were in a noisy environment sometimes, and 24.2 % reported the feeling of being in a noisy environment always. About 51.4 % of teachers reported always having to raise their voice due to presence of noise, while 32.9 % reported having to raise their voice sometimes. Goodman and Kruskal's Gamma showed strong correlations between perception of being in a noisy environment and raising their voices due to it (G=0.876, P<0.001). Also weak to moderate correlations were obtained between raising one's voice and frequency and severity of voice symptoms (P<0.05).

3. *Relationship between voice symptoms and school and classroom location:* Teachers reported their schools to be located close to other schools (60 % teachers), close to government offices and public sectors (23.6 %) and close to quiet streets with residential buildings (16.4 %). Teachers who had their schools located close to other schools, government offices and public sectors experienced more severity of dysphonia. Regarding the classroom location, 55.2 % teachers reported their classrooms to be located close to main traffic roads. A significant correlation was obtained between classroom location and frequency of laryngeal and neck pain that was weekly to monthly recurring (P=0.02).

4. *Relationship between voice symptoms and classroom conditions:* A weak but significant correlation was obtained between severity of throat dryness and classroom size (r=0.18, *P*=0.033). The classrooms accommodated about 36 students on an average. A weak but significant positive correlation was obtained between frequency of dysphonia, that was daily to monthly recurring and the absence of closed doors and windows during teaching hours (G=0.257, *P*=0.024). There were on average four glass windows (non-insulated) and 40 % teachers reported broken wooden doors in their classrooms increasing the perception of noise..

3.1.4.4 Discussion

Teachers tend to raise their voice in presence of noise inside (noise due to student talks and activities and the additive noise intruding from neighboring classrooms) and outside of the classrooms (due to location of schools close to other schools, noisy public sectors and road traffic noise). Due to this, teachers experience increased voice strain, vocal fatigue, throat dryness, frequent throat clearing, neck pain, and dysphonia that recur as a monthly to weekly symptoms and eventually become a daily problem. No voice rest, unawareness of vocal hygiene, and inappropriate use of voice by teachers aggravate the voice problem. Poor classroom conditions (broken doors, no insulated windows, use of fans and power lines increasing low frequency noise) and overcrowded classrooms all elevate the noise in the classrooms causing great annoyance, hindrance and affecting the teachers' general and vocal health. Location of schools near road traffic, overcrowded classrooms, and poor classroom architecture are the biggest source of noise that needs consideration.

In order to overcome these problems, schools in Upper Egypt (and in general) should be designed and constructed in such a way so as to provide a healthy environment for improved performance of teachers and students. Efforts have been made by Egyptian authors to provide guidelines and selection procedures for appropriate planning of school sites, thus considering the geographical, environmental, technical, and safety aspects (Moussa and Elwafa, 2017). Currently, there have also been actions taken to promote 'Green Schools' by the U.S. Green Building Council (USGBC) (Gordon, 2010) in order to construct school buildings that improve teachers' and students' working efficiency. Based on the correlations obtained in our study between teachers' voice symptoms and inappropriate schools and classroom location and conditions, we encourage future studies to measure the internal and external noise levels of not only Egyptian schools, but also schools around the world. Establishing standards for allowable noise levels in and around the schools would further help improved planning of new schools and/or the renovation of existing schools. Such an improved planning in terms of location, infrastructure for improved classroom conditions with optimum acoustical quality similar to those recommended by standards in United States (ANSI/ASA-S12.60, 2009) would be conducive for both teachers' vocal health and the schools would become a better learning place for students.

3.1.4.5 *Conclusion*

The present study demonstrated an influence of noise resulting from the inappropriate location and poor conditions of classrooms and schools on teachers' voices in Upper Egypt. The noise sources reported in this study had a moderate to severe repercussions on the vocal health of our teachers. Therefore, having schools with favorable environmental conditions is necessary to prevent any negative effect on teachers' voice. The present study was based on teachers' self-reports, however, future studies should as well incorporate objective measurements of voice, to have a quantitative data that can be compared prior to and after necessary actions to improve the schools conditions are taken up. Also measuring the schools' internal and external noise is recommended in future.

4 OVERALL CONCLUSION

The present thesis provided insights into various voice assessment procedures. The first part evaluated the vocal fold vibrations using Videokymography and second part assessed teachers' voices with perceptual and acoustic measures and investigating the influence of noise on teachers' voice.

Considering the first part of the thesis on VKG evaluation, in **manuscript I**, VKG was found to be a useful (in 95 % cases) tool in addition to stroboscopy for diagnosis and treatment of voice disorders, especially when the clinician was uncertain of the diagnosis based solely upon stroboscopic evaluation. VKG parameters, particularly the 'rounded lateral peak shape and missing or reduced mucosal waves were found to be helpful in reaching a final diagnosis. There were improvements in the clinician's diagnostic confidence and refinements in treatment recommendations after VKG examination.

In **manuscript II**, the lateral peak sharpness were quantified, and the Open time percentage quotients (OTQ) and Plateau Quotient (PQ) parameters derived at 95 % and 80 % of the vocal fold amplitude, were the most robust objective measures that had strong correlation with visual rating of lateral peak sharpness. These quotients increased their values when the shape of the lateral peak changed from sharp to round.

Considering the second part of this thesis, **manuscript III** brought original data on Cepstral Peak Prominence measures in teachers' voices. The cepstral measure results (CPP and CPPS) obtained in this study could be considered representative for teachers who are functionally healthy and can serve for further comparisons, e.g., with teachers suffering from voice disorders. The thesis also advocates for paying more attention in parameter settings used for cepstral analysis in order to avoid data discrepancies in future studies. The results of **manuscript IV** showed that the noise resulting from inappropriate location and poor conditions of schools and classrooms had a negative impact on teachers' voice in Upper Egypt. This study recommends for a better planning and construction of schools and classrooms in Egypt as well as in schools around the world that will enable better vocal health of teachers. It also encourages future studies to incorporate measures of indoor and outdoor noise levels of schools and classrooms to assess a detailed influence of noise on teachers' voice.

5 REFERENCES

ABO-HASSEBA, A., WAARAMAA, T., ALKU, P. & GENEID, A. 2017. Difference in voice problems and noise reports between teachers of public and private schools in Upper Egypt. *Journal of Voice,* 31**,** 508. e11-508. e16.

ANDRADE MIRANDA, G. X. 2017. *Analyzing of the vocal fold dynamics using laryngeal videos (Doctoral dissertation).* E.T.S.I. Telecomunicación (UPM).

ANGELILLO, I. F., DI MAIO, G., COSTA, G. & BARILLARI, U. 2009. Prevalence of occupational voice disorders in teachers. *Journal of preventive medicine and hygiene,* 50**,** 26-32.

ANSI/ASA-S12.60 2009. American National Standards Institute/Acoustical Society of America. S12.60-2009/Part 2 American National Standard Acoustical Performance Criteria DR, and Guidelines for Schools, Part 2: Relocatable Classroom Factors. Melville, NY: Acoustical Society of America.

ARFFA, R. E., KRISHNA, P., GARTNER-SCHMIDT, J. & ROSEN, C. A. 2012. Normative values for the Voice Handicap Index-10. *Journal of Voice,* 26**,** 462-465.

AWAN, S. N. 2001. *Voice Diagnostic Protocol: A Practical Guide to the Diagnosis of Voice Disorders (Book with CD-ROM for Windows 95 or Higher),* Austin, Texas, PRO-ED, Inc.

AWAN, S. N., GIOVINCO, A. & OWENS, J. 2012. Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of voice,* 26**,** 670. e15-670. e20.

AWAN, S. N., ROY, N., JETTÉ, M. E., MELTZNER, G. S. & HILLMAN, R. E. 2010. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical linguistics & phonetics,* 24**,** 742-758.

BAKEN, R. J. & ORLIKOFF, R. F. 2000. *Clinical measurement of speech and voice,* Singular Publishing Group, San Diego, California.

BAKER, J., BEN-TOVIM, D. I., BUTCHER, A., ESTERMAN, A. & MCLAUGHLIN, K. 2007. Development of a modified diagnostic classification system for voice disorders with inter-rater reliability study. *Logopedics Phoniatrics Vocology,* 32**,** 99-112.

BALASUBRAMANIUM, R. K., BHAT, J. S., FAHIM, S. & RAJU, R. 2011. Cepstral analysis of voice in unilateral adductor vocal fold palsy. *Journal of voice,* 25**,** 326-329.

BERG, F. S., BLAIR, J. C. & BENSON, P. V. 1996. Classroom acoustics: The problem, impact, and solution. *Language, Speech, and Hearing Services in Schools,* 27**,** 16-20.

BEST, S., R.A. & AKAST, L., M. 2016. Visualizing the larynx. In R. T. Sataloff & M. Benninger. *In:* SATALOFF, R. T. (ed.) *Sataloff's Comprehensive Textbook of Otolaryngology: Head & Neck Surgery: Laryngology.* 1st ed. New York: Jaypee Brothers Medical Publishers (P) Ltd.

BLESS, D., PATEL, R. & CONNOR, N. 2009. Laryngeal Imaging: Strobosocpy, High-Speed Digital Imaging, and Kymography. *In:* FRIED, M. & FERLITO, A. (eds.) *The Larynx.* 3 ed.: Plural Publishing, San Diego, CA, Oxford, and Brisbane

BLESS, D. M., HIRANO, M. & FEDER, R. J. 1987. Videostroboscopic evaluation of the larynx. *Ear Nose & Throat Journal,* 66(7)**,** 289-296.

BOERSMA, P. & WEENINK, D. 2013. Praat: Doing phonetics by computer. Amsterdam, the Netherlands: Institute of Phonetic Sciences, University of Amsterdam. http://www.fon.hum.uva.nl/praat/manual/sound_pressure_calibration.html.

BOLBOL, S. A., ZALAT, M. M., HAMMAM, R. A. & ELNAKEB, N. L. 2017. Risk Factors of Voice Disorders and Impact of Vocal Hygiene Awareness Program Among Teachers in Public Schools in Egypt. *Journal of Voice,* 31**,** 251. e9-251. e16.

BONILHA, H. S., DELIYSKI, D. D. & GERLACH, T. T. 2008. Phase asymmetries in normophonic speakers: visual judgments and objective findings. *American Journal of Speech-Language Pathology,* 17**,** 367-376.

BONILHA, H. S., DELIYSKI, D. D., WHITESIDE, J. P. & GERLACH, T. T. 2012. Vocal fold phase asymmetries in patients with voice disorders: A study across visualization techniques. *American Journal of Speech-Language Pathology,* 21**,** 3-15.

BRAUNSCHWEIG, T., FLASCHKA, J., SCHELHORN-NEISE, P. & DOLLINGER, M. 2008. High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias. *Medical Engineering & Physics,* 30**,** 59-66.

BROCKMANN, M., BOHLENDER, J. & MEHTA, D. 2018. Acoustic perturbation measures improve with increasing vocal intensity in individuals with and without voice disorders. *Journal of Voice,* 32**,** 162-168.

CASIANO, R. R., ZAVERI, V. & LUNDY, D. S. 1992. Efficacy of videostroboscopy in the diagnosis of voice disorders. *Otolaryngology-Head and Neck Surgery,* 107**,** 95-100.

CHEN, W., WOO, P. & MURRY, T. 2016. Vocal fold vibratory changes following surgical intervention. *Journal of Voice,* 30**,** 224-227.

CHODARA, A. M., KRAUSERT, C. R. & JIANG, J. J. 2012. Kymographic characterization of vibration in human vocal folds with nodules and polyps. *The Laryngoscope,* 122**,** 58-65.

COLLINS, S. R. 2014. Direct and Indirect Laryngoscopy: Equipment and TechniquesDiscussion. *Respiratory care,* 59**,** 850-864.

COOPER, R. M. 2004. Laryngoscopy—its past and future. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie,* 51**,** R21-R25.

CUTIVA, L. C. C., PUGLISI, G. E., ASTOLFI, A. & CARULLO, A. 2017. Four-day follow-up study on the self-reported voice condition and noise condition of teachers: Relationship between vocal parameters and classroom acoustics. *Journal of Voice,* 31**,** 120.e1-120.e8.

CUTIVA, L. C. C., VOGEL, I. & BURDORF, A. 2013. Voice disorders in teachers and their associations with work-related factors: a systematic review. *Journal of Communication Disorders,* 46**,** 143-155.

CUTLER, J. L. & CLEVELAND, T. 2002. The clinical usefulness of laryngeal videostroboscopy and the role of high-speed cinematography in laryngeal evaluation. *Current Opinion in Otolaryngology & Head and Neck Surgery,* 10**,** 462-466.

CZERMAK, J. 1858. Über den Kehlkopfspiegel. *Wiener Med Wochenschrift,* 7**,** 196-198.

DEJONCKERE, P. H., BRADLEY, P., CLEMENTE, P., CORNUT, G., CREVIER-BUCHMAN, L., FRIEDRICH, G., VAN DE HEYNING, P., REMACLE, M. & 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 Oto-rhino-laryngology,* 258**,** 77-82.

DELIYSKI, D. D. 2010. *Laryngeal high-speed videoendoscopy; in Kendall KA, Leonard RJ, (eds): Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging,* New York, Thieme.

DELIYSKI, D. D. & HILLMAN, R. E. 2010. State of the art laryngeal imaging: research and clinical implications. *Current Opinion in Otolaryngology & Head and Neck Surgery,* 18**,** 147-52.

DURUP, N., SHIELD, B., DANCE, S. & SULLIVAN, R. 2015. An investigation into relationships between classroom acoustic measurements and voice parameters of teachers. *Building Acoustics,* 22**,** 225-241.

DURUP, N., SHIELD, B. M., DANCE, S. & SULLIVAN, R. 2017. Teachers' voice parameters and classroom acoustics—A field study and online survey. *The Journal of the Acoustical Society of America,* 141**,** 3540-3540.

FANT, G. 1960. Acoustic theory of speech perception. *Mouton, The Hague*.

FEX, B., FEX, S., SHIROMOTO, O. & HIRANO, M. 1994. Acoustic analysis of functional dysphonia: before and after voice therapy (accent method). *Journal of Voice,* 8**,** 163-167.

GALL, V., GALL, D. & HANSON, J. 1971. Larynx-fotokymografieLarynx-photokymography. *Archiv für klinische und experimentelle Ohren-Nasen-und Kehlkopfheilkunde,* 200**,** 34-41.

GARCIA, M. 1855. Observations on the human voice. *Proceedings Of The Royal Society Of London* 55**,** 399-410.

GORDON, D. E. 2010. Green Schools as High Performance Learning Facilities *National Clearinghouse for Educational Facilities. https://files.eric.ed.gov/fulltext/ED512700.pdf* [Online].

HAMMARBERG, B., FRITZELL, B., GAUFIN, J., SUNDBERG, J. & WEDIN, L. 1980. Perceptual and acoustic correlates of abnormal voice qualities. *Acta oto-laryngologica,* 90**,** 441-451.
HARTNICK, C. J. & ZEITELS, S. M. 2005. Pediatric video laryngo-stroboscopy. *International journal of pediatric otorhinolaryngology,* 69**,** 215-219.
HASANVAND, A., SALEHI, A. & EBRAHIMIPOUR, M. 2017. A Cepstral Analysis of Normal and Pathologic Voice Qualities in Iranian Adults: A Comparative Study. *Journal of Voice,* 31**,** 508. e17-508. e23.
HEMAN-ACKAH, Y. D., MICHAEL, D. D., BAROODY, M. M., OSTROWSKI, R., HILLENBRAND, J., HEUER, R. J., HORMAN, M. & SATALOFF, R. T. 2003. Cepstral peak prominence: a more reliable measure of dysphonia. *Annals of Otology, Rhinology & Laryngology,* 112**,** 324-333.
HEMAN-ACKAH, Y. D., MICHAEL, D. D. & GODING, G. S. 2002. The Relationship Between Cepstral Peak Prominence and Selected Parameters of Dysphonia. *Journal of Voice,* 16**,** 20-27.
HEMAN-ACKAH, Y. D., SATALOFF, R. T., LAUREYNS, G., LURIE, D., MICHAEL, D. D., HEUER, R., RUBIN, A., ELLER, R., CHANDRAN, S., ABAZA, M., LYONS, K., DIVI, V., LOTT, J., JOHNSON, J. & HILLENBRAND, J. 2014. Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice,* 28**,** 783-788.
HERBST, C. T. & ŠVEC, J. G. 2016. Basics of Voice Acoustics-A Tutorial. . *In:* SATALOFF, R. T. (ed.) *Sataloff's Comprehensive Textbook of Otolaryngology: Head & Neck Surgery: Laryngology.* 1st ed.: Jaypee Brothers Medical Publishers (P) Ltd.
HILLENBRAND, J. 2006. *Version 1.56 [computer program]* [Online].
HILLENBRAND, J., CLEVELAND, R. A. & ERICKSON, R. L. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research,* 37**,** 769-778.
HILLENBRAND, J. & HOUDE, R. A. 1996. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research,* 39**,** 311-321.
HIRANO, M. 1981a. Clinical examination of voice. *Disorders of human communication,* 5**,** 1-99.
HIRANO, M. 1981b. GRBAS scale for evaluating the hoarse voice & frequency range of phonation. *Clinical examination of voice,* 5**,** 83-84.
HIRANO, M. & BLESS, D. M. 1993. *Videostroboscopic examination of the larynx*, Singular Publishing Group San Diego, California.
HIRANO, M., HIBI, S., YOSHIDA, T., HIRADE, Y., KASUYA, H. & KIKUCHI, Y. 1988. Acoustic analysis of pathological voice: some results of clinical application. *Acta oto-laryngologica,* 105**,** 432-438.
HIROTO, I. 1966. The mechanism of phonation; its pathophysiological aspects. *Nippon Jibiinkoka Gakkai Kaiho,* 69**,** 2097-2106.
IEC 2013. IEC 61672-1:2013-Electroacoustics-Sound level meters-Part 1: Specifications *International Electrotechnical Commission.* Geneva, Switzerland.
ILOMÄKI, I., KANKARE, E., TYRMI, J., KLEEMOLA, L. & GENEID, A. 2017. Vocal fatigue symptoms and laryngeal status in relation to vocal activity limitation and participation restriction. *Journal of Voice,* 31**,** 248. e7-248. e10.
ILOMÄKI, I., MÄKI, E. & LAUKKANEN, A.-M. 2005. Vocal symptoms among teachers with and without voice education. *Logopedics Phoniatrics Vocology,* 30**,** 171-174.
ISSHIKI, N., TANABE, M., ISHIZAKA, K. & BROAD, D. 1977. Clinical significance of asymmetrical vocal cord tension. *Annals of Otology, Rhinology & Laryngology,* 86**,** 58-66.
JACOBSON, B. H., JOHNSON, A., GRYWALSKI, C., SILBERGLEIT, A., JACOBSON, G., BENNINGER, M. S. & NEWMAN, C. W. 1997. The voice handicap index (VHI): development and validation. *American Journal of Speech-Language Pathology,* 6**,** 66-70.
JAHN, A. & BLITZER, A. 1996. A short history of laryngoscopy. *Logopedics Phoniatrics Vocology,* 21**,** 181-185.
JÄRVINEN, K., LAUKKANEN, A.-M. & GENEID, A. 2017. Voice quality in native and foreign languages investigated by inverse filtering and perceptual analyses. *Journal of Voice,* 31**,** 261. e25-261. e31.

JIANG, J. J. & TITZE, I. R. 1994. Measurement of vocal fold intraglottal pressure and impact stress. *Journal of Voice,* 8**,** 132-144.

JIANG, J. J., ZHANG, Y., KELLY, M. P., BIEGING, E. T. & HOFFMAN, M. R. 2008. An automatic method to quantify mucosal waves via videokymography. *The Laryngoscope,* 118**,** 1504-1510.

JÓNSDOTTIR, V. I., BOYLE, B. E., MARTIN, P. J. & SIGURDARDOTTIR, G. 2002. A comparison of the occurrence and nature of vocal symptoms in two groups of Icelandic teachers. *Logopedics Phoniatrics Vocology,* 27**,** 98-105.

KANKARE, E., GENEID, A., LAUKKANEN, A.-M. & VILKMAN, E. 2012. Subjective evaluation of voice and working conditions and phoniatric examination in kindergarten teachers. *Folia Phoniatrica et Logopaedica,* 64**,** 12-19.

KASUYA, H., OGAWA, S., KIKUCHI, Y. & EBIHARA, S. 1986. An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology. *Speech Communication,* 5**,** 171-181.

KAY ELEMETRICS 2008. Multi-Dimensional Voice Program, Model 5105 Lincoln Park, NJ: Kay Elemetrics Corporation.

KEMPSTER, G. B., GERRATT, B. R., ABBOTT, K. V., BARKMEIER-KRAEMER, J. & HILLMAN, R. E. 2009. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology,* 18**,** 124-132.

KENDALL, K. A. 2009. High-speed laryngeal imaging compared with videostroboscopy in healthy subjects. *Archives of Otolaryngology–Head & Neck Surgery,* 135**,** 274-281.

KENT, R. D. & BALL, M. J. 2000. *Voice quality measurement,* Singular Publishing Group San Diego.

KENT, R. D. & READ, C. 1992. *The acoustic analysis of speech,* Singular Publishing Group San Diego.

KIM, G., LEE, Y., PARK, H., BAE, I. & KWON, S. 2017. A study of cepstral peak prominence characteristics in ADSV, SpeechTool and Praat. *Journal of Speech-Language & Hearing Disorders,* 26**,** 99-111.

KOUFMAN & ISAACSON, G. 1991. The spectrum of vocal dysfunction. *Otolaryngologic Clinics of North America,* 24**,** 985-988.

KRISTIANSEN, J., LUND, S. P., PERSSON, R., SHIBUYA, H., NIELSEN, P. M. & SCHOLZ, M. 2014. A study of classroom acoustics and school teachers' noise exposure, voice load and speaking time during teaching, and the effects on vocal and mental fatigue development. *International archives of occupational and environmental health,* 87**,** 851-860.

KUMAR, B. R., BHAT, J. S. & PRASAD, N. 2010. Cepstral analysis of voice in persons with vocal nodules. *Journal of Voice,* 24**,** 651-653.

KUMAR, S. P., PHADKE, K. V., VYDROVÁ, J., NOVOZÁMSKÝ, A., ZITA, A., ZITOVÁ, B. & ŠVEC, J. G. 2018. Visual and Automatic Evaluation of Vocal Fold Mucosal Waves Through Sharpness of Lateral Peaks in High-Speed Videokymographic Images. *Journal of Voice (In Press)*.

LATOSZEK, B. B. V., DE BODT, M., GERRITS, E. & MARYN, Y. 2018. The exploration of an objective model for roughness with several acoustic markers. *Journal of Voice,* 32**,** 149-161.

LAW, T., KIM, J. H., LEE, K. Y., TANG, E. C., LAM, J. H., VAN HASSELT, A. C. & TONG, M. C. 2012. Comparison of rater's reliability on perceptual evaluation of different types of voice sample. *Journal of Voice,* 26**,** 666. e13-666. e21.

LOPES, L. W., DA SILVA, J. D., SIMÕES, L. B., DA SILVA EVANGELISTA, D., SILVA, P. O. C., ALMEIDA, A. A. & DE LIMA-SILVA, M. F. B. 2017. Relationship between acoustic measurements and self-evaluation in patients with voice disorders. *Journal of Voice,* 31**,** 119. e1-119. e10.

MADILL, C., NGUYEN, D. D., EASTWOOD, C., HEARD, R. & WARHURST, S. 2018. Comparison of Cepstral Peak Prominence Measures Using the ADSV, SpeechTool, and VoiceSauce Acoustic Analysis Programs in Vocally Healthy Female Speakers. *Acoustics Australia,* 46**,** 215-226.

MARYN, Y. & ROY, N. 2012. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. *Jornal da Sociedade Brasileira de Fonoaudiologia,* 24**,** 107-112.

MARYN, Y., ROY, N., DE BODT, M., VAN CAUWENBERGE, P. & CORTHALS, P. 2009. Acoustic measurement of overall voice quality: a meta-analysis. *The Journal of the Acoustical Society of America,* 126**,** 2619-2634.

MARYN, Y. & WEENINK, D. 2015. Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice,* 29**,** 35-43.

MCDONALD, KATIE & ERIK, T. 2011. Cepstral peak prominence as a method for gauging ethnic differences in phonation. *Paper presented at New Ways of Analysing Variation 40, 27 October.* Washington DC.

MCGOWAN, R. 1990. An analogy between the mucosal waves of the vocal folds and wind waves on water. *Haskins Laboratories Status Report on Speech Research,* 101**,** 243-9.

MEHTA, D. D., DELIYSKI, D. D. & HILLMAN, R. E. 2010a. Commentary on why laryngeal stroboscopy really works: Clarifying misconceptions surrounding Talbot's law and the persistence of vision. *Journal of Speech, Language, and Hearing Research,* 53**,** 1263-1267.

MEHTA, D. D., DELIYSKI, D. D., ZEITELS, S. M., QUATIERI, T. F. & HILLMAN, R. E. 2010b. Voice production mechanisms following phonosurgical treatment of early glottic cancer. *Annals of Otology, Rhinology & Laryngology,* 119**,** 1-9.

MEHTA, D. D. & HILLMAN, R. E. 2012. Current role of stroboscopy in laryngeal imaging. *Current opinion in otolaryngology & head and neck surgery,* 20**,** 429.

MEHTA, D. D., ZAÑARTU, M., QUATIERI, T. F., DELIYSKI, D. D. & HILLMAN, R. E. 2011. Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy. *the Journal of the Acoustical Society of America,* 130**,** 3999-4009.

MENDELSOHN, A. H., REMACLE, M., COUREY, M. S., GERHARD, F. & POSTMA, G. N. 2013. The diagnostic role of high-speed vocal fold vibratory imaging. *Journal of Voice,* 27**,** 627-31.

MORAWSKA, J. & NIEBUDEK-BOGUSZ, E. 2017. Risk factors and prevalence of voice disorders in different occupational groups-a review of literature. *Otorynolaryngologia,* 16**,** 94-102.

MORROW, S. L. & CONNOR, N. P. 2011. Comparison of voice-use profiles between elementary classroom and music teachers. *Journal of Voice,* 25**,** 367-372.

MOUSSA, M. & ELWAFA, A. A. 2017. School site selection process. *Procedia Environmental Sciences,* 37**,** 282-293.

MUNIER, C. & KINSELLA, R. 2007. The prevalence and impact of voice problems in primary school teachers. *Occupational Medicine,* 58**,** 74-76.

NAWKA, T., ANDERS, L. C. & WENDLER, J. 1994. Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache Stimme Gehör,* 18**,** 130-133.

NELSON, P. 1999. The Changing Demand for Improved Acoustics in Our Schools. *Volta Review,* 101**,** 23-31.

NIEBUDEK-BOGUSZ, E., KUZANSKA, A., WOZNICKA, E. & SLIWINSKA-KOWALSKA, M. 2011. Assessment of the voice handicap index as a screening tool in dysphonic patients. *Folia phoniatrica et logopaedica,* 63**,** 269-72.

NIEDZIELSKA, G. 2001. Acoustic analysis in the diagnosis of voice disorders in children. *International journal of pediatric otorhinolaryngology,* 57**,** 189-193.

NOLL, A. M. 1964. Short time spectrum and "cepstrum" techniques for vocal pitch detection. *The Journal of the Acoustical Society of America,* 36**,** 296-302.

NOVOZÁMSKÝ, A., SEDLÁŘ, J., ZITA, A., ŠROUBEK, F., FLUSSEF, J., ŠVEC, J. G., VYDROVÁ, J. & ZITOVÁ, B. Image analysis of videokymographic data. Image Processing (ICIP), 2015 IEEE International Conference on, 2015. IEEE, 78-82.

OATES, J. 2009. Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatrica et Logopaedica,* 61**,** 49-56.

OERTEL, M. 1895. Das Laryngo-Stroboskop die Laryngostroboscopische Untersuchung. *Arch Laryngol Rhinol,* 3**,** 1-16.

OERTEL, M. 1878. Über eine neue "laryngostroboskopische" Untersuchungsmethode des Kehlkopfes. *Zbl Med Wissensch,* 16**,** 81-82.

OLTHOFF, A., WOYWOD, C. & KRUSE, E. 2007. Stroboscopy versus high-speed glottography: a comparative study. *The Laryngoscope,* 117**,** 1123-1126.

PATEL, R., AWAN, S. N., BARKMEIER-KRAEMER, J., COUREY, M., DELIYSKI, D., EADIE, T., PAUL, D., ŠVEC, J. G. & HILLMAN, R. 2018. Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American journal of speech-language pathology***,** 1-19.

PATEL, R., DAILEY, S. & BLESS, D. 2008. Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Annals of Otology, Rhinology & Laryngology,* 117**,** 413-424.

PAUL, B. C., CHEN, S., SRIDHARAN, S., FANG, Y., AMIN, M. R. & BRANSKI, R. C. 2013. Diagnostic accuracy of history, laryngoscopy, and stroboscopy. *The Laryngoscope,* 123**,** 215-219.

PIAZZA, C., MANGILI, S., DEL BON, F., GRITTI, F., MANFREDI, C., NICOLAI, P. & PERETTI, G. 2012. Quantitative analysis of videokymography in normal and pathological vocal folds: a preliminary study. *European archives of oto-rhino-laryngology,* 269**,** 207-212.

PICKUP, B. & THOMSON, S. 2009. Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models. *Journal of biomechanics,* 42**,** 2219-2225.

PRINTZA, A., TRIARIDIS, S., THEMELIS, C. & CONSTANTINIDIS, J. 2012. Stroboscopy for benign laryngeal pathology in evidence based health care. *Hippokratia,* 16**,** 324.

PUGLISI, G. E., ASTOLFI, A., CANTOR CUTIVA, L. C. & CARULLO, A. 2017. Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. *The Journal of the Acoustical Society of America,* 141**,** 441-452.

QIU, Q. & SCHUTTE, H. K. 2007. Real-time kymographic imaging for visualizing human vocal-fold vibratory function. *Review of Scientific Instruments,* 78**,** 024302.

RANTALA, L. M., HAKALA, S., HOLMQVIST, S. & SALA, E. 2015. Classroom noise and teachers' voice production. *Journal of Speech, Language, and Hearing Research,* 58**,** 1397-1406.

ROGERSON, J. & DODD, B. 2005. Is there an effect of dysphonic teachers' voices on children's processing of spoken language? *Journal of Voice,* 19**,** 47-60.

ROSEN, C. A., AMIN, M. R., SULICA, L., SIMPSON, C. B., MERATI, A. L., COUREY, M. S., JOHNS, M. M., 3RD & POSTMA, G. N. 2009. Advances in office-based diagnosis and treatment in laryngology. *Laryngoscope,* 119 Suppl 2**,** S185-212.

ROSEN, C. A., GARTNER-SCHMIDT, J., HATHAWAY, B., SIMPSON, C. B., POSTMA, G. N., COUREY, M. & SATALOFF, R. T. 2012. A nomenclature paradigm for benign midmembranous vocal fold lesions. *The Laryngoscope,* 122**,** 1335-1341.

ROSEN, C. A., LEE, A. S., OSBORNE, J., ZULLO, T. & MURRY, T. 2004. Development and validation of the voice handicap index-10. *The Laryngoscope,* 114**,** 1549-1556.

ROY, N., MERRILL, R. M., THIBEAULT, S., GRAY, S. D. & SMITH, E. M. 2004. Voice disorders in teachers and the general population: effects on work performance, attendance, and future career choices. *Journal of Speech, Language, and Hearing Research,* 47**,** 542-551.

SALA, E., LAINE, A., SIMBERG, S., PENTTI, J. & SUONPÄÄ, J. 2001. The prevalence of voice disorders among day care center teachers compared with nurses: a questionnaire and clinical study. *Journal of Voice,* 15**,** 413-423.

SALDANHA, D., KRISHNAMOORTHY, V., HENSEL, P., MUELLER, M., LAURITO, C. I. & WEINBERG, G. 2013. Preoperative indirect mirror laryngoscopy used as means of predicting a difficult airway. *Journal of Clinical Anesthesia,* 25**,** 250.

SATALOFF, R. T., GULLANE, P. J. & GOLDSTEIN, D. P. 2015. *Sataloff's Comprehensive Textbook of Otolaryngology: Head & Neck Surgery: Head and Neck Surgery,* JP Medical Ltd.

SATALOFF, R. T., SPIEGEL, J. R. & HAWKSHAW, M. J. 1991. Strobovideolaryngoscopy: results and clinical value. *Annals of Otology, Rhinology & Laryngology,* 100**,** 725-727.
SAUDER, C., BRETL, M. & EADIE, T. 2017. Predicting voice disorder status from smoothed measures of cepstral peak prominence using Praat and analysis of dysphonia in speech and voice (ADSV). *Journal of Voice,* 31**,** 557-566.
SCHWARTZ, S. R., COHEN, S. M., DAILEY, S. H., ROSENFELD, R. M., DEUTSCH, E. S., GILLESPIE, M. B., GRANIERI, E., HAPNER, E. R., KIMBALL, C. 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, J. P., COWLEY, T., MCCOY, S., BERNAD, P. G. & PATEL, M. M. 2009. Clinical practice guideline: hoarseness (dysphonia). *Otolaryngol Head Neck Surg,* 141**,** S1-S31.
SHAW, H. S. & DELIYSKI, D. D. 2008. Mucosal wave: a normophonic study across visualization techniques. *Journal of Voice,* 22**,** 23-33.
SHUE, Y.-L., CHEN, G. & ALWAN, A. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. Eleventh Annual Conference of the International Speech Communication Association, INTERSPEECH, 34-37 https://www.isca-speech.org/archive/interspeech_2010/i10_0034.html, 2010.
SILBERMAN, H. D., WILF, H. & TUCKER, J. A. 1976. Flexible fiberoptic nasopharyngolaryngoscope. *Annals of Otology, Rhinology & Laryngology,* 85**,** 640-645.
SIMBERG, S., SALA, E., VEHMAS, K. & LAINE, A. 2005. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *Journal of Voice,* 19**,** 95-102.
SLIWINSKA-KOWALSKA, M., NIEBUDEK-BOGUSZ, E., FISZER, M., LOS-SPYCHALSKA, T., KOTYLO, P., SZNUROWSKA-PRZYGOCKA, B. & MODRZEWSKA, M. 2006. The prevalence and risk factors for occupational voice disorders in teachers. *Folia Phoniatrica et Logopaedica,* 58**,** 85-101.
ŠRAM, F., ŠVEC, J. G. & VYDROVA, J. 2018. Videokymography. In: Zehn- hoff-Dinnesen A, Wiskirska-Woznica B, Neumann KE. *Nawka T (eds) European manual of medicine: phoniatrics***,** 1. Springer, Berlin (In Press)
STORY, B. H. 2002. An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology,* 23**,** 195-206.
SULICA, L. 2011. Hoarseness. *Archives of Otolaryngology–Head & Neck Surgery,* 137**,** 616-619.
SULTER, A. & ALBERS, F. 1996. The effects of frequency and intensity level on glottal closure in normal subjects. *Clinical Otolaryngology & Allied Sciences,* 21**,** 324-327.
ŠVEC, J., FRIČ, M., ŠRAM, F. & SCHUTTE, H. K. Mucosal Waves on the Vocal Folds: Conceptualization Based on Videokymography. Fifth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, 2007a.
ŠVEC, J. & ŠRAM, F. 2011. Videokymographic examination of voice. *In:* MA, E. P. M. & YIU, E. M. L. (eds.) *Handbook of voice assessments.* Plural Publishing, San Diego, CA.
ŠVEC, J., ŠRAM, F. & SCHUTTE, H. 2009. Videokymography. *In:* FRIED, M. & FERLITO, A. (eds.) *The larynx.* 3 ed.: Plural Publishing,San Diego, CA.
ŠVEC, J. G. & GRANQVIST, S. 2010. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology,* 19**,** 356-368.
ŠVEC, J. G. & GRANQVIST, S. 2018. Tutorial and Guidelines on Measurement of Sound Pressure Level in Voice and Speech. *Journal of Speech, Language, and Hearing Research,* 61**,** 441-461.
ŠVEC, J. G. & SCHUTTE, H. K. 1996. Videokymography: high-speed line scanning of vocal fold vibration. *Journal of Voice,* 10**,** 201-205.
ŠVEC, J. G. & SCHUTTE, H. K. 2012. Kymographic imaging of laryngeal vibrations. *Current opinion in otolaryngology & head and neck surgery,* 20**,** 458-465.
ŠVEC, J. G., ŠRAM, F. & SCHUTTE, H. K. 2007b. Videokymography in voice disorders: what to look for? *Annals of Otology, Rhinology & Laryngology,* 116**,** 172-180.

ŠVEC, J. G., ŠVECOVÁ, H., HERBST, C. & SCHUTTE, H. K. 2007c. Evaluation protocol for videokymographic images. (MsAccess software application). Groningen, the Netherlands: Groningen Voice Research Lab, University of Groningen
TEPPO, H. & ALHO, O. P. 2008. Relative importance of diagnostic delays in different head and neck cancers. *Clinical Otolaryngology,* 33**,** 325-330.
THIBEAULT, S. L., MERRILL, R. M., ROY, N., GRAY, S. D. & SMITH, E. M. 2004. Occupational risk factors associated with voice disorders among teachers. *Annals of epidemiology,* 14**,** 786-792.
TIESLER, G. & OBERDÖRSTER, M. 2008. Noise—a stressor? Acoustic ergonomics of schools. *Building Acoustics,* 15**,** 249-261.
TITZE, I. R. 1988. The physics of small-amplitude oscillation of the vocal folds. *The Journal of the Acoustical Society of America,* 83**,** 1536-1552.
TITZE, I. R. 1995. *Workshop on acoustic voice analysis: Summary statement*, National Center for Voice and Speech.
TITZE, I. R. 2000. Principles of voice production (second printing). Iowa City, IA, National Center for Voice and Speech.
TITZE, I. R. 2001. Acoustic interpretation of resonant voice. *Journal of voice,* 15**,** 519-528.
TITZE, I. R., JIANG, J. J. & HSIAO, T.-Y. 1993. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. *Annals of Otology, Rhinology & Laryngology,* 102**,** 58-63.
TITZE, I. R. & LIANG, H. 1993. Comparison of Fo extraction methods for high-precision voice perturbation measurements. *Journal of Speech, Language, and Hearing Research,* 36**,** 1120-1133.
VERDOLINI, K., ROSEN, C. & BRANSKI, R. 2006. American speech-language-hearing association special interest division 3, voice and voice disorders, classification manual for voice disorders–I. Mahwah, NJ: Lawrence Erlbaum Associates.
VILKMAN, E. 2000. Voice problems at work: a challenge for occupational safety and health arrangement. *Folia phoniatrica et logopaedica,* 52**,** 120-125.
WATTS, C. R. & AWAN, S. N. 2011. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research,* 54**,** 1525-1537.
WATTS, C. R., AWAN, S. N. & MARYN, Y. 2017. A comparison of cepstral peak prominence measures from two acoustic analysis programs. *Journal of Voice,* 31**,** 387. e1-387. e10.
WHO. 2016. *ICD-10 Version: 2016* [Online]. http://apps.who.int/classifications/icd10/browse/2016/en.
WOLFE, V. & MARTIN, D. 1997. Acoustic correlates of dysphonia: type and severity. *Journal of Communication Disorders,* 30**,** 403-416.
WOO, P. 2016. Stroboscopy and high-speed video examination of the larynx. *In:* SATALOFF, R. T. & BENNINGER, M. (eds.) *Sataloff's Comprehensive Textbook of Otolaryngology: Head & Neck Surgery: Laryngology.* New York: Jaypee Brothers Medical Publishers (P) Ltd.
WOO, P., COLTON, R., CASPER, J. & BREWER, D. 1991. Diagnostic value of stroboscopic examination in hoarse patients. *Journal of voice,* 5**,** 231-238.
YAMAUCHI, A., YOKONISHI, H., IMAGAWA, H., SAKAKIBARA, K.-I., NITO, T. & TAYAMA, N. 2015b. Quantitative Analysis of Vocal Fold Vibration in Vocal Fold Paralysis With the Use of High-speed Digital Imaging. *Journal of Voice,* 30**,** 766.e13-766.e22
YAMAUCHI, A., YOKONISHI, H., IMAGAWA, H., SAKAKIBARA, K.-I., NITO, T., TAYAMA, N. & YAMASOBA, T. 2015. Quantitative analysis of digital videokymography: a preliminary study on age-and gender-related difference of vocal fold vibration in normal speakers. *Journal of Voice,* 29**,** 109-119.
YAMAUCHI, A., YOKONISHI, H., IMAGAWA, H., SAKAKIBARA, K.-I., NITO, T., TAYAMA, N. & YAMASOBA, T. 2016a. Visualization and estimation of vibratory disturbance in vocal fold scar using high-speed digital imaging. *Journal of Voice,* 30**,** 493-500.

YAMAUCHI, A., YOKONISHI, H., IMAGAWA, H., SAKAKIBARA, K.-I., NITO, T., TAYAMA, N. & YAMASOBA, T. 2016b. Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging. *Journal of Voice,* 30**,** 205-214.

YIU, E. M. 1999. Limitations of perturbation measures in clinical acoustic voice analysis. *Asia Pacific Journal of Speech, Language and Hearing,* 4**,** 155-166.

ZACHARIAS, S. R., DELIYSKI, D. D. & GERLACH, T. T. 2018. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *Journal of Voice,* 32**,** 216-220.

ZHANG, Y. & JIANG, J. J. 2008. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *Journal of Voice,* 22**,** 1-9.

ZHANG, Z. 2016. Mechanics of human voice production and control. *The Journal of the Acoustical Society of America,* 140**,** 2614-2635.

ZHANG, Z. & HIEU LUU, T. 2012. Asymmetric vibration in a two-layer vocal fold model with left-right stiffness asymmetry: Experiment and simulation. *The Journal of the Acoustical Society of America,* 132**,** 1626-1635.

ZIEGER, K., SCHNEIDER, C., GERULL, G. & MROWINSKI, D. 1995. Cepstrum analysis in voice disorders. *Folia phoniatrica et logopaedica,* 47**,** 210-217.

ZUR, K. B., COTTON, S., KELCHNER, L., BAKER, S., WEINRICH, B. & LEE, L. 2007. Pediatric Voice Handicap Index (pVHI): A new tool for evaluating pediatric dysphonia. *International Journal of Pediatric Otorhinolaryngology,* 71**,** 77-82.

6 MANUSCRIPT SUPPLEMENTS

Supplement A: Manuscript I

Supplement B: Manuscript II

Supplement C: Manuscript III

Supplement D: Manuscript IV

Supplement A: Manuscript I

Evaluation of Clinical Value of Videokymography for Diagnosis and Treatment of Voice Disorders

Authors and their contribution to the study in percentages:

K. V. Phadke 50 %
J. Vydrová 15 %
R. Domagalská 10 %
J. G. Švec 25 %

Journal name and Impact factor (IF):

European Archives of Oto-Rhino-Laryngology, IF (2017): 1.546

Authors' contribution:

K. V. Phadke: Contributed to the design of the study, designed the questionnaire used in the study, analyzed the data, wrote the manuscript.

J. Vydrová: Contributed to the design of the questionnaire, examined the patients, filled the questionnaires, checked and corrected the final version of the manuscript.

R. Domagalská: Organized and managed the paperwork with patients and the videorecordings, organized filling-in the questionnaires, provided the patient data.

J. G Švec: Initiated, designed and supervised the study, critically revised the successive versions of the manuscript.

Eur Arch Otorhinolaryngol (2017) 274:3941–3949
DOI 10.1007/s00405-017-4726-1

LARYNGOLOGY

Evaluation of clinical value of videokymography for diagnosis and treatment of voice disorders

Ketaki Vasant Phadke[1] · Jitka Vydrová[2] · Romana Domagalská[2] · Jan G. Švec[1]

Received: 13 April 2017 / Accepted: 21 August 2017 / Published online: 30 August 2017

Abstract This study aimed at determining the clinical value of videokymography (VKG) as an additional tool for the assessment of voice disorders. 105 subjects with voice disorders were examined by an experienced laryngologist. A questionnaire was used to specify diagnosis, diagnostic confidence, and treatment recommendations before and after VKG. The first part of questionnaire was filled by the laryngologist for each patient after routine ear-nose-throat evaluation, including stroboscopy, the second part after the subsequent VKG examination. In 31% of subjects VKG confirmed the stroboscopic diagnosis, in 44% it made the diagnosis more accurate, in 20% there was adjustment of the treatment, and in 5% it was not found diagnostically useful. After VKG the diagnostic confidence increased in 68% of the subjects. VKG may help clinicians to take some important treatment decisions and may be recommended to be performed in patients, where clinicians are uncertain about diagnosis and treatment.

Keywords Stroboscopy · Videokymography · Clinical value · Vocal folds · Vibration characteristics

Electronic supplementary material The online version of this article (doi:10.1007/s00405-017-4726-1) contains supplementary material, which is available to authorized users.

✉ Jan G. Švec
Jan.Svec@upol.cz

[1] Voice Research Lab, Department of Biophysics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

[2] Voice and Hearing Centre Prague, Medical Healthcom Ltd, Španělská 4, Prague, Czech Republic

Introduction

Stroboscopy has become important for the diagnosis and treatment decisions of various voice disorders and it is considered as a gold standard instrument for assessing various voice problems [1, 2]. Several studies have been carried out to quantify the diagnostic value of stroboscopy in assessing various voice disorders by comparing it with the diagnosis made based on previous traditional visualization techniques such as mirror and fiberoptic examination of larynx [3], subjective impressions and laryngoscopy under continuous light [4], indirect laryngoscopy [5], and history with physical examination and flexible laryngoscopy [6]. Stroboscopy has been reported to have confirmed or changed the initial diagnosis and treatment decisions made based on these traditional visualization techniques, thus proving its clinical usefulness in assessing various voice disorders [3–8]. In spite of the limitation of videostroboscopy in its sampling frequency, it provides a real-time visualization of vocal fold vibration at slow motion; and thus remains voice clinician's preferred choice of imaging modality.

However, since stroboscopy is confined to only periodic vibration of vocal folds, it hampers the diagnostic possibilities in patients with hoarse voice quality. High-speed imaging techniques, therefore, have been increasingly used as they allow assessment of vocal fold behavior also when stroboscopy fails, such as in cases of short or aperiodic vocal fold vibrations due to its advantage of high sampling rate [9]. It has been suggested that high-speed videoendoscopy (HSV) may be used as an alternative to stroboscopy especially for the diagnosis of functional voice disorders, since HSV allows assessing phonation onset, irregularities, and all aperiodic patterns of vocal fold vibration [10]. Others augment HSV to stroboscopy in cases when it is difficult to use such as in moderate to severe dysphonia [2].

Some authors have also found HSV to be advantageous in examining voice production mechanism in individuals with postsurgical early glottic cancer [11] and quantitatively analyzing vibratory changes after vocal fold surgery [12]. However, HSV requires more advanced objective analysis procedures, for better understanding of the vocal fold movement patterns [13]. Such analysis procedures are rather time consuming. The other disadvantages of HSV include the lack of immediate visual feedback, higher cost of the equipment, often inferior image quality compared to videostroboscopy (due to lower spatial resolution in relation to increased recording speed), and large amount of data requiring higher storage space and making HSV computationally more demanding [14–16]. Moreover, there is a lack of evidence on the clinical relevance of HSV, as there is no standard clinical protocol established so far for use of HSV in functional assessment of voice disorders [17].

In the last decades, videokymography (VKG) has emerged as a simpler and more economical alternative to HSV [14, 16, 18]. It is a single-line scanning high-speed imaging technique with a focus on assessing various voice disorders based on vocal fold vibration characteristics. Numerous types of vibratory characteristics were identified for sustained phonations (presence or absence of vocal fold vibration; interference of surrounding structures; cycle-to-cycle variability; closure duration; opening versus closing duration; left–right asymmetry; shape of the lateral peaks; shape of the medial peaks; mucosal waves and cycle aberrations) providing information on different types of voice problems [19]. Since VKG is a real-time imaging technique, it allows for immediate clinical feedback, that is, the examiner is able to visualize the kymographic images simultaneously during the examination. It has a high spatial resolution (over 700 pixels/line) and high image rate (7200 line images/second) [20].

While VKG has been reported to be helpful in better recognizing some types of voice disorders [19, 21, 22], its diagnostic value and clinical relevance has not yet been formally evaluated. Although some clinicians use VKG routinely, many others are not sure how this technique could be useful. There has been no study addressing the usefulness of VKG for the diagnosis and treatment decisions and the types of voice disorders in which VKG could yield useful information. Hence, this study was designed to evaluate the added clinical value of VKG to stroboscopy in evaluation of individuals with various voice disorders. The specific questions were as follows: (1) Does VKG help in establishing the diagnosis of voice disorders in patients? (2) Does VKG improve the clinician's confidence in diagnosis based on stroboscopy? (3) Does VKG help in treatment decisions?

Materials and methods

Participants

105 individuals (71 females and 34 males; aged between 10 and 80 years; 42 of them were professional voice users) served as subjects for the study. They were prospectively selected from the patients coming to the department to be examined for voice problems. They primarily complained of hoarseness, vocal fatigue, and loss of vocal range. All the participants underwent a detailed case history and routine introductory ear-nose-throat (ENT) procedure followed by stroboscopic and VKG examination of voice. The VKG examination was indicated for (1) individuals with voice complaints, having normal findings on stroboscopy, (2) for singers with voice problems, (3) for verification of vibration problems in case of structural abnormality (hemorrhages, sulcus, leukoplakia, suspicion of tumor, vocal fold edema and laryngitis), (4) for individuals with organic findings with unclear influences on vocal fold vibrations. VKG was not indicated for individuals with acute infections, subjects with clear organic findings and with innervation disorders in which the vocal fold vibration problems were secondary. The first two columns of Table 1 list the types of voice disorders presented by the participants included in the study and their distribution.

Instrumentation

The stroboscopic examination was done using the EndoSTROB workplace including a Matrix LED duo light source and a 90° rigid laryngoscope with an integrated chip-on-the-tip video camera (all Xion, Berlin, Germany). For the VKG examination, the second generation VKG camera (Kymocam, CYMO, b.v. Groningen, the Netherlands, image rate 7200 lines/s) was connected to the laryngoscope (Xion Medical, Germany, 10 mm diameter, 90° angle) using a C-mount objective adapter (R. Wolf, Germany, type 85261.272, 27 mm focal length). The larynx was illuminated by a 300 W endoscopic xenon light source (type FX 300 A, Fentex Medical, Germany). Both the stroboscopic and videokymographic recordings were stored digitally by means of the EndoSTROB video capturing unit.

Questionnaire

For the purpose of investigating the clinical value of VKG a questionnaire was designed, which was divided into two parts, first half focusing on the clinical value and diagnosis confidence from stroboscopic evaluation and the second half on the clinical contribution of VKG as an addition to stroboscopy (available as a supplementary material). In the stroboscopic part, the clinician filled in the diagnosis according

Table 1 Distribution of diagnoses of study participants and of the ratings for diagnostic usefulness of VKG

Diagnosis	Cases	Distribution of ratings of usefulness (cases)				
		$R = 0$	$R = 1$	$R = 2$	$R = 3$	$R = 4$
Laryngitis (chronic and acute, incl. reflux)	65	3	16	29	17	–
Hyperfunctional dysphonia	10	1	5	4	–	–
Vocal fatigue	5	–	2	1	2	–
Vocal fold paralysis	4	–	2	2	–	–
Vocal polyp	4	–	–	4	–	–
Larynx and vocal fold edema	4	–	1	3	–	–
Singing technique problems with normal laryngeal findings	3	–	3	–	–	–
Nasopharyngitis with normal laryngeal findings	2	–	1	–	1	–
Vocal nodules	2	–	–	1	1	–
Mutational voice disorder	2	–	1	1	–	–
Dilated blood vessels	1	–	–	1	–	–
Vocal fold carcinoma	1	–	1	–	–	–
Spasmodic dysphonia	1	1	–	–	–	–
Asthma bronchiole and coughing exacerbation	1	–	1	–	–	–
Total (no.)	105	5	33	46	21	0
Total (%)	100	5	31	44	20	0

Rating categories: no diagnostic contribution of VKG ($R = 0$); VKG confirmed the stroboscopic diagnosis ($R = 1$); VKG made the diagnosis more accurate ($R = 2$); VKG made the diagnosis more accurate and resulted in an adjustment of offered treatment ($R = 3$); and VKG changed the initial diagnosis and changed the offered treatment ($R = 4$)

to the International Classification of Diseases (ICD10 Version: 2016) and provided with additional diagnostic information, if needed. The confidence with the diagnosis was rated on a five-point scale (0—not confident; 1—little confident; 2—moderately confident; 3—greatly confident; 4—absolutely confident). The questionnaire also included treatment recommendations after the stroboscopic evaluation.

The second, VKG part of the questionnaire consisted of specifying the final diagnosis after stroboscopy and VKG examination, rating the confidence with the final diagnosis, specifying the treatment recommendation, listing the VKG features which were found helpful for the diagnosis, and rating the usefulness of VKG on a five-point scale (0—no diagnostic contribution of VKG; 1—VKG confirmed the stroboscopic diagnosis; 2—VKG made the diagnosis more accurate; 3—VKG made the diagnosis more accurate and resulted in an adjustment of offered treatment; and 4—VKG changed the initial diagnosis and changed the offered treatment).

Evaluation procedure

Before stroboscopy, the clients underwent routine introductory ENT procedure, including patient history, ear, nose, and oral cavity examination. For the stroboscopy exam, the participants were seated comfortably and the clinician inserted the rigid laryngoscope into the subject's mouth while he/she was instructed to sustain a phonation of vowel /i/ at comfortable pitch and loudness. In case of need, the subject was instructed to vary pitch and loudness in order to better elucidate the problem. After the strobe exam the clinician filled in the first part of the questionnaire. Following this, the same individuals underwent a VKG examination in a similar fashion as done previously in stroboscopy and the clinician filled in the second part of the questionnaire. The examination and filling in the questionnaires were done by a single clinician (a senior laryngologist of the Center with 30 years of experience in ENT practice and over 10 years of experience of working with VKG) with the help of a nurse and junior laryngologists. The stroboscopic and VKG evaluation procedures were identical to those performed at the site in routine clinical practice; no procedures were added here for the purpose of the present study, except of the filling in of the questionnaire. All stroboscopic and VKG evaluations were done based on visual observation, and no extensive quantitative analysis was performed by the clinician due to time constraints in routine clinical setup.

Data analysis

The data from all the questionnaires were transferred to an MS Excel spread sheet for further analysis. Origin 2015 (OriginLab Corporation, Northampton, USA) software was used for analysis of results and preparation of graphs. Descriptive statistics was used to determine the diagnostic value of VKG, clinician's confidence in diagnosis and the most important VKG features that helped for final diagnosis.

Results

The results of the rating of the clinical value of VKG in addition to stroboscopy from 105 cases are shown in Fig. 1. In 95% of cases, VKG was found useful for establishing the diagnosis. Within these cases, in 31% individuals VKG confirmed the initial stroboscopic diagnosis, in 44% individuals VKG made the diagnosis more accurate, and in 20% individuals VKG evaluation resulted in an adjustment of offered treatment. In 5% of individuals VKG was rated not useful. In no case the VKG evaluation was found critical by completely changing the stroboscopic diagnosis and treatment. The usefulness ratings for the different diagnoses can be found in Table 1.

The diagnostic confidence of the clinician after the initial stroboscopic evaluation and after the added VKG examination is shown in Fig. 2. When performing stroboscopic examination, the clinicians reported being greatly confident in their diagnosis in 53% individuals and absolutely confident in 8% individuals. After VKG was performed the level of diagnostic confidence shifted to higher values. The clinicians were greatly confident with their diagnosis in 42% and absolutely confident in 51% individuals. Overall, the diagnostic confidence level increased in 68% cases and remained the same in 32% cases after the clinicians performed the VKG examinations.

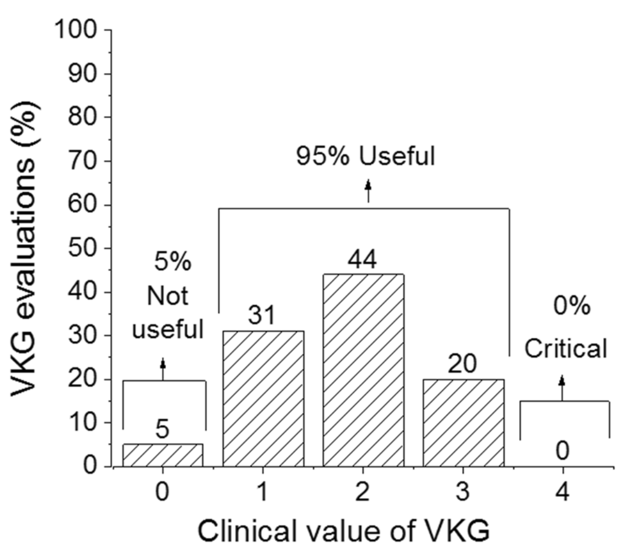


Fig. 1 The clinical value of videokymography (VKG) in 105 evaluations expressed on a five-point scale (0—no diagnostic contribution of VKG; 1—VKG confirmed the stroboscopic diagnosis; 2—VKG made the diagnosis more accurate; 3—VKG made the diagnosis more accurate and resulted in an adjustment of offered treatment; and 4—VKG changed the initial diagnosis and changed the offered treatment)

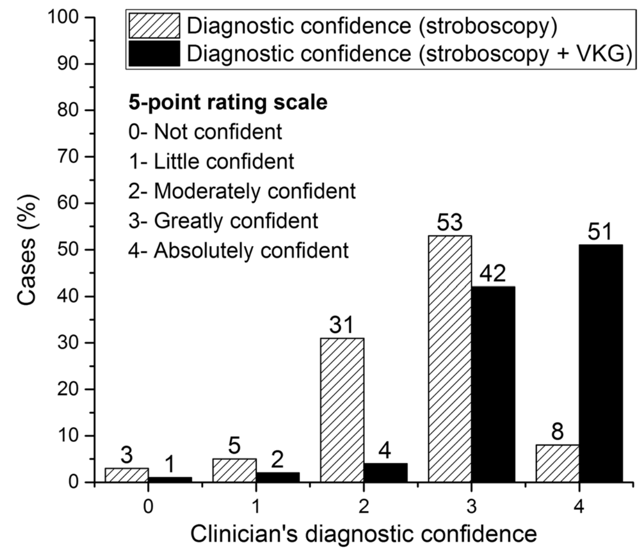


Fig. 2 The clinician's diagnostic confidence after stroboscopy and after the VKG examination for the 105 cases

Discussion

While videostroboscopy is considered to have a principal clinical role in laryngeal imaging [23], recently there have been interrogations pertaining to the perceptual rating of stroboscopic parameters, particularly associated with glottal closure, phase closure, phase asymmetry, and irregularity of vocal fold vibrations, leading to poor reliability and questionable overall validity of stroboscopic rating [24]. On the contrary, VKG was reported to be more reliable in assessing these parameters as it is known to reliably record irregular vibrations [19]. Also, it allows assessing the vocal fold vibration features in a single kymographic image, containing essential vibratory information, usually missed on videostroboscopy examination [14]. The use of kymographic images obtained digitally from HSV recordings and digital kymograms (DKGs) was reported to allow better assessment of vibratory irregularities (glottal width and period irregularities) in normophonic speakers than using video recordings from HSV and videostroboscopy [25]. Kymography was also found to be superior in measuring the vibratory asymmetries of the vocal folds than videostroboscopy and full view HSV recordings [26–28].

Clinical value of videokymography

Referring to Fig. 1, and Table 1, VKG was found clinically useful in 95% individuals. These were mostly subjects initially diagnosed with acute and chronic laryngitis (including laryngopharyngeal reflux disease) and hyperfunctional dysphonia. Included were also subjects with smaller lesions, vocal fatigue, and voice disorders due to other systemic conditions such as upper respiratory infections. Out of the 95% useful cases, VKG confirmed the initial stroboscopic diagnosis in 31% cases, made the diagnosis more accurate in 44% cases, and resulted in altered treatment in 20% cases. Here, it should be noted that VKG was primarily expected not to provide structural diagnosis (polyp, cyst, etc.) but rather bring more insights into the seriousness of the voice disorder based on the nature of alterations from normal vibratory behavior using the empirical rule: "the more abnormal the vocal fold vibration, the more serious the voice disorder". Normal vocal fold vibratory behavior was recognized based on the following VKG features [16, 29]: (a) both vocal folds are vibrating, (b) ventricular folds are not vibrating, (c) vibrational amplitudes of both vocal folds are approximately similar, (d) vibrational frequencies of both vocal folds are approximately the same, (e) the vibrations are regular, (f) The vibrations are free of aberrations, (g) the vocal folds touch each other during vibration at the place of maximum vibration amplitude, (h) the closed phase takes between ca. 10 and 60% of period duration at the place of maximum vibration amplitude, (i) the shape of lateral peaks (i.e., the turn from opening to closing) is sharp, (j) mucosal waves propagate laterally on the upper vocal fold surface, and (k) no large left–right phase asymmetry is present. In the cases of VKG confirming the diagnosis, the vibratory behavior of the vocal fold was as expected, without further insights into the nature or degree of the disorder. In the cases of VKG making the stroboscopic diagnosis more accurate, the extent of deviation of the VKG features from their normal visual appearance was used as an indicator of the extent of the pathology and allowed refining the judgment on how much the pathology influences the vibratory behavior of the vocal fold. In cases of nodules, the VKG features (particularly presence of vibration, mucosal waves and shape of lateral peaks) were used to assess whether the nodules were pliable and thus whether they were at an early stage of development. Similarly, the VKG features were used to assess how much an infection affected the pliability of vocal fold mucosa.

Figure 3 shows a case example in which VKG was found useful in making the diagnosis more accurate. A female subject (48 years, hair dresser) complained of persistent hoarseness for the last 2 months and intermittent hoarseness lasting for few years. Originally, she thought the hoarseness to be a result of infection but routine anti-inflammatory treatments including antibiotics were not effective. Stroboscopy showed accumulating mucus and slight swelling in the mid-membranous portion of the vocal folds. The membranous glottal closure was complete. There were signs of slight mucosal damage and reddening of arytenoids, leading to suspicion of laryngitis caused by laryngopharyngeal reflux disorder (LPRD). VKG showed clear pathological alterations in vocal fold vibration, with rounded lateral peaks, absent mucosal wave on left and reduced mucosal wave extent on right vocal fold. As these vibratory abnormalities are strong signs of

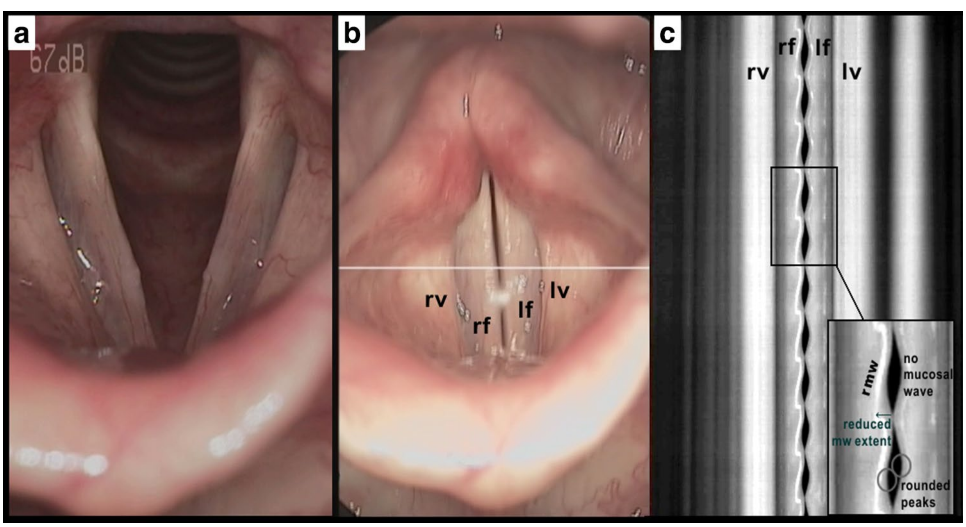


Fig. 3 Images from a subject complaining of persistent hoarseness for last 2 months. **a** Laryngoscopic image of open vocal folds during breathing, **b** Standard laryngoscopic image from the VKG camera during phonation showing the position of the scan line at the middle of glottis used for VKG examination. **c** The VKG image (40 ms duration, time running down) from the position indicated in (**b**) showing the vibratory behavior of the left (lf) and right vocal folds (rf) and the nonvibrating ventricular folds (lv, rv). The magnified snapshot shows rounded lateral peaks on both vocal folds and missing mucosal wave on the left fold indicating stiffened mucosa. On the right fold the mucosal wave is present (rmw) but its extent is shortened to less than half of the vocal fold width, also indicating slight stiffening of the mucosa

excessive stiffness of the vocal fold mucosa [16, 20], they made the original stroboscopic diagnosis of laryngitis more accurate. They also increased clinician's confidence in the original diagnosis and justified the need for LPRD testing. Finally, LPRD was diagnosed using esophageal 24-h multichannel intraluminal impedance pH monitoring. Therapy included proton pump inhibitors (PPIs), dietary, and lifestyle recommendations.

In the cases of VKG altering the treatment offered, the severity of the VKG findings led to refinements in voice therapy (4% subjects), change in medication (3%), modification in dosage of drugs (10%), and clarification in duration of voice rest (3%). Figure 4 shows a case example in which VKG made the diagnosis more accurate and resulted in an adjustment of offered treatment. A female professional singer (age 27 years) complained of reduced sonority of her singing voice, vocal fatigue, phlegm, and slight coughing together with headaches and enervation. She had an important concert ahead and asked whether she could sing without endangering her vocal folds. Her throat, tonsils, and pharynx showed no signs of infection. Epipharyngoscopy discovered reddened nasal mucosa and phlegm in nasopharyngeal arcs. Ultrasonography revealed largely thickened mucosa in maxillary sinuses with a free center, the frontal sinuses were aerial. Laryngoscopy showed free laryngeal entrance, normal epiglottis as well as normal ventricular and aryepiglottic folds. Stroboscopy showed pale vocal folds with slightly ectatic vessels vibrating along the whole glottal length, glottal closure was complete along the whole membranous length; there was only some mucus accumulating at the place of maximum vibration amplitude. The tracheal mucosa was reddened. VKG nevertheless showed normal vibratory features of the vocal folds along the whole glottal length in low, medium, and high pitches. Because of the normal VKG features the clinician allowed the singer to hold the concert. The final diagnosis was nasopharyngitis and sinusitis of mucosal type. The singer was prescribed with antibiotics (ATB) and decongestants to suppress the inflammation in nasopharyngeal and maxillary cavities. The ATB did not prevent her from doing singing exercises and continue concert rehearsals. On the third day, she went through the concert successfully and finished taking the ATB on the seventh day. VKG played an important role here as it showed normal vibratory behavior of the vocal folds and thus helped to diagnose no tissue damage of the vocal folds during the infection allowing the singer to perform the concert.

VKG was rated not useful (rating of 0) in 5% individuals. These were five subjects initially diagnosed based on stroboscopy as having dysphonia, spasmodic dysphonia, chronic reflux laryngitis, acute laryngitis, and chronic laryngitis with atrophy of left vocal fold, respectively. The VKG did not indicate presence of any other pathology that would otherwise be missed on stroboscopy, and therefore the VKG was rated not useful in these individuals.

There were no cases (0%) in this study in which the VKG evaluation was found critical by completely changing the stroboscopic diagnosis and treatment decision. However, there have been reports in which HSV and videokymography

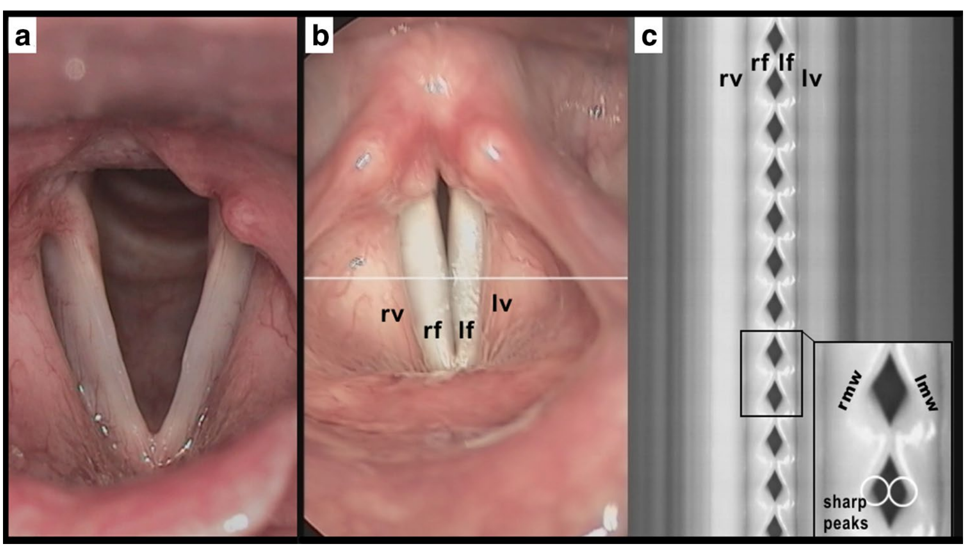


Fig. 4 Images from a subject (professional singer) complaining of worsened singing voice due to an infection of upper breathing airways. **a** Laryngoscopic image of open vocal folds during breathing. **b** Standard laryngoscopic image from the VKG camera during phonation showing the position of the scan line at the middle of glottis used for VKG examination. **c** VKG image (40 ms duration, time running down) showing normal vibration features on the left and right vocal folds (lf, rf) and the nonvibrating ventricular folds (lv, rv). The magnified snapshot shows the presence of both left and right mucosal waves (lmw, rmw), sharp lateral peaks on both vocal folds [encircled in the zoomed image in (**c**)], left–right symmetry in phase, amplitude and frequency of vibration, and presence of glottal closure

enabled discovering vocal fold carcinoma previously missed on stroboscopy [30, 31]. Therefore, although such a case did not occur within the pool of subjects for this study, it should also be kept in mind and the diagnostic value of VKG should be explored in more detail on a larger group of voice disorders.

Clinician's diagnostic confidence

VKG examination rendered clinicians with additional information regarding the vibratory behaviors of the vocal fold. This improved the clinician's diagnostic confidence when they were uncertain about their initial diagnosis based on stroboscopy. Most of the subjects in this study had symptoms of hoarseness and vocal fatigue. When evaluating hoarseness, it is essential for clinicians to evaluate the alterations of vibrations due to vocal fold lesion or asymmetry [32]. Stroboscopy is problematic in capturing some abnormal features (cycle-to-cycle irregularities, roundedness of lateral vibratory peaks, vibratory asymmetries, etc.) [14, 16, 19]. In such circumstances, VKG may be a helpful tool in providing additional information about abnormality of the vocal fold vibration, thus helping clinicians to better diagnose the voice disorders. VKG allowed becoming more confident in the diagnosis also in cases of normal laryngeal findings in stroboscopy. These appeared in five subjects. Three out of the five subjects were singers who came for a control evaluation, with problems in singing high notes. VKG confirmed normal vibratory abilities of their vocal folds and indicated problems with singing technique rather than organic findings. In the remaining two subjects, VKG increased diagnostic confidence by confirming normal vocal fold behavior in nasopharyngitis.

Helpful VKG features

The occurrence of VKG features that were listed as helpful for the final diagnosis is depicted in Table 2. The most frequently occurring VKG features in our subjects were rounded lateral peaks (58%) and reduced/absent mucosal waves (44%). Both these features are reported to be related to reduced mucosal pliability [14, 33, 34]. The mucosal waves traveling upwards along the medial surface of the vocal fold cause vertical phase difference between the movement of the lower and upper margins, which are then reflected as sharp lateral peaks in the VKG images [14, 19, 34]. Reduced vertical phase differences cause the lateral peaks to be rounded rather than sharp. Rounded lateral peaks are thus a sign of reduced mucosal pliability which is related to impaired vocalization due to reduced delivery of energy from the airflow to the vocal folds [14, 16, 18, 19]. Rounded lateral peaks are well visible in VKG whereas they are difficult to observe in stroboscopy. As such, this feature can help in better recognizing mucosal pliability impairments in inflammatory conditions such as laryngitis or in stiff localized lesions such as nodules or polyps [35]. Evaluation of mucosal pliability plays a vital role when assessing voice disorders [16, 36]. Specific changes in the vocal folds due to injury result in an associated change in the mucosal wave properties [37–39]. Mucosal wave properties reflect vocal competence [37, 40], health, and pliability of mucosa of the vocal fold [14].

Table 2 The occurrence of VKG features that were listed as helpful for the diagnosis in the 105 cases

VKG features	Cases (%)
1. Rounded lateral peaks	58
2. Missing or reduced mucosal wave	44
3. Left–right phase differences	27
4. Reduced vocal fold amplitude	26
5. Missing glottal closure	16
6. Normal VKG findings	15
7. Missing vocal fold vibration	8
8. Irregular vocal fold vibration	5
9. Glottal closure too short	3
10. Sharp medial peaks	3
11. Left–right frequency differences	2
12. Interference of vibrating ventricular folds and surrounding structures	2

Other VKG features which were reported helpful were vocal fold asymmetry in terms of left–right phase difference (27%) and amplitude differences (26%). These differences are observed when there is a left–right asymmetry of structure, tension, or mass of the vocal folds [40, 41]. These features were relatively common in our subjects diagnosed with nodules, individuals with chronic laryngitis, or vocal fold paralysis. However, left–right asymmetry has been known to occur also in normal individuals [25, 42–49], so the presence of left–right phase asymmetry is not expected to be the only feature to decide on the diagnosis of voice disorder.

In about 15% of the individuals, VKG showed normal findings. These included individuals treated for chronic laryngitis, vocal fatigue, etc. Normal VKG findings in these individuals were considered a positive sign indicating that the disorder did not have a severe effect on the vocal fold vibration. This helped the clinicians to decide on when to terminate the treatment or, in case of professional singers, whether the singer could be allowed to perform on stage.

Among other useful VKG features there were missing glottal closure (16%), missing (8%) and irregular vocal fold vibration (5%). Other VKG features including glottal closure too short (3%), sharp medial peaks (3%), ventricular vocal fold vibrations (2%) were less frequently occurring in the subjects evaluated in this study and their usefulness needs to be further evaluated on larger variety of voice disorders.

The results of this study are confined to limited types of voice disorder based on their occurrence in daily clinical practice of our institute. For example, the study consisted of 62% subjects diagnosed with some type of laryngitis and only 1% with vocal fold cancer. Some diagnoses, in which vibrational behavior was considered as unnecessary to evaluate for final diagnosis and treatment (such as acute virosis, papilloma, unilateral vocal fold paralysis, clear organic lesions such as nodules and polyps, etc.), were excluded due to time constraints imposed by the daily clinical practice. Since different institutes could have different distribution of diagnoses and different time constraints, VKG role may vary among these. Other limitation of this study is that it is based mostly on assessments of a single clinician experienced with using VKG. Adding more clinicians would increase the strength of the study, but they were not available due to personal and time constraints at the institute. Future studies should address these limitations and also bring more information on the correspondence and complementarity of the videostroboscopic and videokymographic findings.

Conclusion

Videokymography was found diagnostically useful in 95% of the cases where it confirmed the initial stroboscopic diagnosis (31%), made the diagnosis more accurate (44%), or resulted in adjustment of the treatment (20%). Furthermore, the clinician's confidence in the diagnosis improved after the VKG examination in 68% of the cases. VKG offered information on vibration properties which was supplementary to stroboscopic findings. Particularly, the information on the sharpness/roundedness of the lateral peaks and on the presence and extent of mucosal wave offered by VKG was frequently used to evaluate the pliability of mucosa, inflammatory tissue infiltration, and lesion stiffness. This helped the clinician to take some important treatment decisions pertaining to changing medication, modifying drug dosage, refining or ending voice therapy, and clarifying duration of voice rest period. VKG examination was found useful in patients when the clinician was uncertain about their diagnosis and treatment decisions.

Acknowledgements The authors would like to thank J. Dubová, MD and E. Vitásková, MD for their help in collecting the data.

Compliance with ethical standards

Ethical approval The study was approved by the Ethics Committee of the Voice and Hearing Centre Prague, Medical Healthcom, Ltd. The subjects have signed a consent form approving using their data anonymously for the study.

Funding The study was supported by the Technology Agency of the Czech Republic project no. TA04010877.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Hartnick CJ, Zeitels SM (2005) Pediatric video laryngo-stroboscopy. Int J Pediatr Otorhinolaryngol 69(2):215–219
2. Patel R, Dailey S, Bless D (2008) Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. Ann Otol Rhinol Laryngol 117(6):413–424
3. Woo P, Colton R, Casper J, Brewer D (1991) Diagnostic value of stroboscopic examination in hoarse patients. J Voice 5(3):231–238. doi:10.1016/S0892-1997(05)80191-2
4. Sataloff RT, Spiegel JR, Hawkshaw MJ (1991) Strobovideolaryngoscopy: results and clinical value. Ann Otol Rhinol Laryngol 100(9):725–727. doi:10.1177/000348949110000907
5. Casiano RR, Zaveri V, Lundy DS (1992) Efficacy of videostroboscopy in the diagnosis of voice disorders. Otolaryngol Head Neck Surg 107(1):95–100
6. Paul BC, Chen S, Sridharan S, Fang Y, Amin MR, Branski RC (2013) Diagnostic accuracy of history, laryngoscopy, and stroboscopy. Laryngoscope 123(1):215–219
7. Printza A, Triaridis S, Themelis C, Constantinidis J (2012) Stroboscopy for benign laryngeal pathology in evidence based health care. Hippokratia 16(4):324
8. Sataloff RT, Spiegel JR, Carroll LM, Schiebel BR, Darby KS, Rulnick R (1988) Strobovideolaryngoscopy in professional voice users: results and clinical value. J Voice 1(4):359–364. doi:10.1016/S0892-1997(88)80012-2
9. Rosen CA, Amin MR, Sulica L et al (2009) Advances in office-based diagnosis and treatment in laryngology. Laryngoscope 119(Suppl 2):S185–S212
10. Braunschweig T, Flaschka J, Schelhorn-Neise P, Dollinger M (2008) High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias. Med Eng Phys 30(1):59–66
11. Mehta DD, Deliyski DD, Zeitels SM, Quatieri TF, Hillman RE (2010) Voice production mechanisms following phonosurgical treatment of early glottic cancer. Ann Otol Rhinol Laryngol 119(1):1–9
12. Chen W, Woo P, Murry T (2016) Vocal fold vibratory changes following surgical intervention. J Voice 30(2):224–227
13. Olthoff A, Woywod C, Kruse E (2007) Stroboscopy versus high-speed glottography: a comparative study. Laryngoscope 117(6):1123–1126
14. Svec J, Sram F, Schutte H (2009) Videokymography. In: Fried M, Ferlito A (eds) The larynx, vol 1, 3rd edn. Plural Publishing, San Diego, pp 253–271
15. Bless D, Patel R, Connor N (2009) Laryngeal imaging: strobosocpy, high-speed digital imaging, and kymography. In: Fried M, Ferlito A (eds) The larynx, 3rd edn. Plural Publishing, San Diego, pp 182–191
16. Svec J, Sram F (2011) Videokymographic examination of voice. In: Ma EPM, Yiu EML (eds) Handbook of voice assessments. Plural Publishing, San Diego, pp 129–146
17. Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE (2007) Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. Folia Phoniatrica et Logopaedica 60(1):33–44

18. Svec JG, Schutte HK (1996) Videokymography: high-speed line scanning of vocal fold vibration. J Voice 10(2):201–205
19. Svec JG, Sram F, Schutte HK (2007) Videokymography in voice disorders: what to look for? Ann Otol Rhinol Laryngol 116(3):172–180
20. Svec JG, Schutte HK (2012) Kymographic imaging of laryngeal vibrations. Curr Opin Otolaryngol Head Neck Surg 20(6):458–465
21. Piazza C, Mangili S, Del Bon F et al (2012) Quantitative analysis of videokymography in normal and pathological vocal folds: a preliminary study. Eur Arch Otorhinolaryngol 269(1):207–212
22. Chodara AM, Krausert CR, Jiang JJ (2012) Kymographic characterization of vibration in human vocal folds with nodules and polyps. Laryngoscope 122(1):58–65
23. Mehta DD, Hillman RE (2012) Current role of stroboscopy in laryngeal imaging. Curr Opin Otolaryngol Head Neck Surg 20(6):429
24. Nawka T, Konerding U (2012) The interrater reliability of stroboscopy evaluations. J Voice 26(6):812.e1–812.e10. doi:10.1016/j.jvoice.2011.09.009
25. Bonilha HS, Deliyski DD (2008) Period and glottal width irregularities in vocally normal speakers. J Voice 22(6):699–708
26. Bonilha HS, Deliyski DD, Whiteside JP, Gerlach TT (2012) Vocal fold phase asymmetries in patients with voice disorders: a study across visualization techniques. Am J Speech Lang Pathol 21(1):3–15
27. Mehta DD, Deliyski DD, Quatieri TF, Hillman RE (2011) Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings. J Speech Lang Hear Res 54(1):47–54
28. Yamauchi A, Yokonishi H, Imagawa H et al (2015) Vocal fold vibration in vocal fold atrophy: quantitative analysis with high-speed digital imaging. J Voice 29(6):755–762. doi:10.1016/j.jvoice.2014.12.008
29. Sram F, Svec JG, Vydrova J (2018) Videokymography. In: Zehnhoff-Dinnesen A, Wiskirska-Woznica B, Neumann K, Nawka T (eds) European manual of medicine: phoniatrics 1. Springer, Berlin **(in press)**
30. Vydrova J, Svec J, Sram F (2015) Videokymography (VKG) in laryngologic practice. J MacroTrends Health Med 3:87–95. http://macrojournals.com/yahoo_site_admin/assets/docs/8HM31Vr.32215729.pdf
31. Unger J, Lohscheller J, Reiter M, Eder K, Betz CS, Schuster M (2015) A noninvasive procedure for early-stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis. Can Res 75(1):31–39
32. Kendall KA, Browning MM, Skovlund SM (2005) Introduction to high-speed imaging of the larynx. Curr Opin Otolaryngol Head Neck Surg 13(3):135–137
33. Yamauchi A, Yokonishi H, Imagawa H et al (2016) Visualization and estimation of vibratory disturbance in vocal fold scar using high-speed digital imaging. J Voice 30(4):493–500. doi:10.1016/j.jvoice.2015.07.003
34. Sundberg J, Högset C (2001) Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. Logop Phoniatr Vocology 26(1):26–36
35. Yamauchi A, Yokonishi H, Imagawa H et al (2016) Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging. J Voice 30(2):205–214. doi:10.1016/j.jvoice.2015.04.016
36. Jiang JJ, Zhang Y, Kelly MP, Bieging ET, Hoffman MR (2008) An automatic method to quantify mucosal waves via videokymography. Laryngoscope 118(8):1504–1510
37. Jiang JJ, Chang CI, Raviv JR, Gupta S, Banzali FM, Hanson DG (2000) Quantitative study of mucosal wave via videokymography in canine larynges. Laryngoscope 110(9):1567–1573
38. Li L, Zhang Y, Maytag AL, Jiang JJ (2015) Quantitative study for the surface dehydration of vocal folds based on high-speed imaging. J Voice 29(4):403–409. doi:10.1016/j.jvoice.2014.09.025
39. Yamauchi A, Yokonishi H, Imagawa H et al (2017) Characterization of vocal fold vibration in sulcus vocalis using high-speed digital imaging. J Speech Lang Hear Res 60(1):24–37
40. Yamauchi A, Yokonishi H, Imagawa H, Sakakibara KI, Nito T, Tayama N (2016) Quantitative analysis of vocal fold vibration in vocal fold paralysis with the use of high-speed digital imaging. J Voice 30(6):766.e13–766.e22. doi:10.1016/j.jvoice.2015.10.015
41. Isshiki N, Tanabe M, Ishizaka K, Broad D (1977) Clinical significance of asymmetrical vocal cord tension. Ann Otol Rhinol Laryngol 86(1):58–66
42. Elias ME, Sataloff RT, Rosen DC, Heuer RJ, Spiegel JR (1997) Normal strobovideolaryngoscopy: variability in healthy singers. J Voice 11(1):104–107
43. Haben CM, Kost K, Papagiannis G (2003) Lateral phase mucosal wave asymmetries in the clinical voice laboratory. J Voice 17(1):3–11
44. Kendall KA (2009) High-speed laryngeal imaging compared with videostroboscopy in healthy subjects. Arch Otolaryngol Head Neck Surg 135(3):274–281
45. Krenmayr A, Wöllner T, Supper N, Zorowka P (2012) Visualizing phase relations of the vocal folds by means of high-speed videoendoscopy. J Voice 26(4):471–479
46. Lacina O (1970) Die adduktionelle Asymmetrie des Kehlkopfes bei den Sängern (asymmetria arytaenoidea cruciata cantatorum). Folia Phoniatrica et Logopaedica 22(2):100–106
47. Lindestad P-Å, Hertegård S, Björck G (2004) Laryngeal adduction asymmetries in normal speaking subjects. Logop Phoniatr Vocology 29(3):128–134
48. Lohscheller J, Švec JG, Döllinger M (2013) Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects. Logop Phoniatr Vocology 38(4):182–192
49. Yamauchi A, Yokonishi H, Imagawa H et al (2015) Quantitative analysis of digital videokymography: a preliminary study on age- and gender-related difference of vocal fold vibration in normal speakers. J Voice 29(1):109–119

Supplementary material to the article

Title: Evaluation of Clinical Value of Videokymography for Diagnosis and Treatment of Voice Disorders

Journal name: European Archives of Oto-Rhino-Laryngology and Head & Neck

Authors: Ketaki.Vasant.Phadke[1], Jitka Vydrová[2], Romana. Domagalská [2], Jan G. Švec [1]

Affiliation:

[1]Department of Biophysics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

[2]Voice and Hearing Centre Prague, Medical Healthcom Ltd, Španělská 4, Prague, Czech Republic

Corresponding author information:

Jan G. Švec, Ph.D. et Ph.D.

Palacký University Olomouc, Faculty of Science, Dept. Biophysics, Voice Research Lab, 17. Listopadu 12, 771 46 Olomouc, Czech Republic

Phone Number: +420 58 563 4151, Email Address: Jan.Svec@upol.cz

Questionnaire

CLINICAL VALUE OF STROBOSCOPY AND VIDEOKYMOGRAPHY:

Patient name**:**

Date of birth or ID number**:**

Date of examination**:**

Physician**:**

Present complaints**:**

Brief history of voice problems:

Previous diagnoses, health-related factors**:**

STROBOSCOPY:

Diagnosis based on stroboscopy:

Fill in the diagnosis in words:

ICD 10 Classification:

How confident are you with your diagnosis made from stroboscopy?

(please tick appropriate)

0- Not confident

1- Little confident

2- Moderately confident

3- Greatly confident

4- Absolutely confident

Treatment recommendation based on stroboscopy:

(Please select, add)

a) Antibiotic treatment:

b) Antiedematic treatment:

c) Antireflux treatment:

d) Voice therapy: which?

e) Surgical treatment: which?

f) Others:which?

VIDEOKYMOGRAPHY (VKG):

Diagnosis based on VKG:

Fill in the diagnosis in words:

ICD 10 Classification:

How confident are you with your diagnosis made from both the VKG and stroboscopic recordings?

(please tick appropriate)

0- Not confident

1- Little confident

2- Moderately confident

3- Greatly confident

4- Absolutely confident

Which VKG feature was important for VKG diagnosis?

(please tick appropriate)

a) Normal VKG
b) Missing VF vibration
c) Reduced VF amplitude
d) Missing or reduced mucosal wave
e) Rounded lateral peaks
f) Sharp medial peaks
g) Glottal closure missing
h) Glottal closure too short
i) Glottal closure to long
j) Irregular VF vibrations
k) Frequency differences
l) Phase differences
m) Ventricular folds vibrating
n) Other feature. Please specify which one:

Treatment recommendation based on VKG:

(Please select, add)

a) Antibiotic treatment:

b) Antiedematic treatment:

c) Antireflux treatment:

d) Voice therapy: which?

e) Surgical treatment: which?

f) Others: which?

Clinical value of VKG additional to stroboscopy:

(please tick appropriate)

0 - No diagnostic contribution of VKG

1 -VKG confirmed the stroboscopic diagnosis

2 -VKG made the diagnosis more accurate

3 -VKG made the diagnosis more accurate and resulted in an adjustment of offered treatment (explain how)

4 -VKG changed the initial diagnosis and changed the offered treatment (explain how)

Supplement B: Manuscript II

Visual and Automatic Evaluation of Vocal Fold Mucosal Waves Through Sharpness of Lateral Peaks in High-Speed Videokymographic Images

Authors and their contribution to the study in percentages:

S.P. Kumar	40 %
K. V. Phadke	**15 %**
J. Vydrová	5 %
A. Novozámský	5 %
A. Zita	5 %
B. Zitová	5 %
J.G Švec	25 %

Journal name and Impact factor (IF):

Journal of Voice, IF (2016): 1.381

Authors' contribution:

S.P. Kumar: Performed visual ratings of the sharpness of lateral peak from videokymographic images, designed automatic method for quantifying the sharpness of lateral peaks from vocal fold contours, analyzed the data and wrote the manuscript.

K. V. Phadke: Extracted the videokymographic images from patient recordings that represented different lateral peak shapes; extracted the vocal fold contours using VKG analyzer software, performed visual rating of the images; contributed to data analysis and statistical analysis of results, participated in writing the manuscript.

J. Vydrová: Examined the patients using videokymography, corrected the final version of the manuscript.

A. Novozámský: Developed the VKG analyzer software for extracting vocal fold contours, corrected the final version of the manuscript.

A. Zita: Developed the VKG analyzer software for extracting vocal fold contours, investigated the peak sharpness in VKG images, checked the final version of the manuscript.

B. Zitová: Supervised the development of the VKG analyzer software, checked and revised the final version of the manuscript.

J.G. Švec: Initiated, designed and supervised the study and the data analysis, performed visual ratings of the images, critically revised the successive versions of the manuscript.

Visual and Automatic Evaluation of Vocal Fold Mucosal Waves Through Sharpness of Lateral Peaks in High-Speed Videokymographic Images

*S. Pravin Kumar, *Ketaki Vasant Phadke, †Jitka Vydrová, ‡Adam Novozámský, ‡Aleš Zita, ‡Barbara Zitová, and *Jan G. Švec, *Olomouc, and †,‡Prague, Czech Republic

Abstract: **Introduction.** The sharpness of lateral peaks is a visually helpful clinical feature in high-speed videokymographic (VKG) images indicating vertical phase differences and mucosal waves on the vibrating vocal folds and giving insights into the health and pliability of vocal fold mucosa. This study aims at investigating parameters that can be helpful in objectively quantifying the lateral peak sharpness from the VKG images.

Method. Forty-five clinical VKG images with different degrees of sharpness of lateral peaks were independently evaluated visually by three raters. The ratings were compared to parameters obtained by automatic image analysis of the vocal fold contours: *Open Time Percentage Quotients* (OTQ) and *Plateau Quotients* (PQ). The OTQ parameters were derived as fractions of the period during which the vocal fold displacement exceeds a predetermined percentage of the vibratory amplitude. The PQ parameters were derived similarly but as a fraction of the open phase instead of a period.

Results. The best correspondence between the visual ratings and the automatically derived quotients were found for the OTQ and PQ parameters derived at 95% and 80% of the amplitude, named OTQ_{95}, PQ_{95}, OTQ_{80} and PQ_{80}. Their Spearman's rank correlation coefficients were in the range of 0.73 to 0.77 ($P < 0.001$) indicating strong relationships with the visual ratings. The strengths of these correlations were similar to those found from inter-rater comparisons of visual evaluations of peak sharpness.

Conclusion. The Open time percentage and Plateau quotients at 95% and 80% of the amplitude stood out as the possible candidates for capturing the sharpness of the lateral peaks with their reliability comparable to that of visual ratings.

Keywords: Mucosal waves−Lateral peak sharpness−Kymography−Vocal fold vibration−Image analysis−Quantification.

INTRODUCTION

The occurrence of mucosal waves on the vibrating vocal folds has been generally recognized as a crucial indicator for healthy voice. Mucosal waves originate at the inferior surface of the vocal fold mucosa, propagate vertically along the medial surface, and then horizontally along the superior surface, creating a wave-like motion on the vocal folds.[1-8] A soft and pliable superficial layer of the lamina propria is necessary for their occurrence.[1,9] In other words, health and pliability of vocal fold mucosa may be indicated by the presence of mucosal waves.[10] Reduced mucosal wave amplitude is clinically observed in cases of increased mucosal stiffness due to, eg, lesions or scarring.[9,11]

Observations on excised hemilarynges[5,7,12-15] and lately also ultrasonic laryngeal observations *in vivo*[16] have shown that mucosal waves are associated with the phase-delayed movements of the upper vocal fold margin (lip or edge) trailing the lower margin. This delay is termed “vertical phase difference”, and it facilitates the delivery of airflow energy to vocal fold tissue.[2,3,10,17-19] Titze et al (1993)[5] stroboscopically tracked the flesh-points in excised larynges to quantify the phase delay and demonstrated its relationship with mucosal wave propagation velocity.

In vivo laryngoscopic imaging techniques such as videostroboscopy and high-speed videoendoscopy (HSV) have enabled easier visualization and quantitative evaluation of the presence, absence, or reduction of mucosal waves in clinical practice.[1,20-28] An alternative view for clinical evaluation of the mucosal waves has been offered by kymographic (ie, single-line) imaging techniques such as videokymography (VKG), digital kymography (DKG) or strobovideokymography (SVKG).[29]

Kymography assesses mucosal waves based on (1) vertical phase differences and (2) laterally traveling mucosal waves.[10,30,31] Vertical phase differences show up as sharp lateral peaks in kymograms, and laterally running mucosal

Accepted for publication August 30, 2018.

A part of the work was presented at the Voice Foundation's 47th Annual Symposium − Care of the Professional Voice, May 30-June 3, 2018, Philadelphia, USA. At this symposium S.P.K. received the Hamdan International Presenter Award 2018 and Sataloff New Investigator Award 2018 for this work. The research of the authors was supported by Technological Agency of the Czech Republic project no. TA04010877 (2014-2017) and the Czech Science Foundation (GA CR) project no. GA16-01246S (2016-2018). S. P. K. thanks Sri Sivasubramaniya Nadar (SSN) College of Engineering, Kalavakkam, India, for granting the leave of absence during his sabbatical stay (2016-2018) at Palacký University in Olomouc, Czech Republic.

From the *Voice Research Lab, Department of Biophysics, Faculty of Science, Palacký University, Olomouc, Czech Republic; †Voice and Hearing Centre, Medical Healthcom Ltd., Prague, Czech Republic; and the ‡Department of Image Processing, Institute of Information Theory and Automation of the Czech Academy of Sciences, Prague, Czech Republic.

Address correspondence and reprint requests to S. Pravin Kumar or Jan G. Švec, Voice Research Lab, Department of Biophysics, Faculty of Science, Palacký University, 17. listopadu 12, Olomouc 77416, Czech Republic. E-mail: pravin.subbaraj@upol.cz; jan.svec@upol.cz

Journal of Voice, Vol. ■■, No. ■■, pp. 1−9

0892-1997

https://doi.org/10.1016/j.jvoice.2018.08.022

waves appear on the kymogram as lines running obliquely sidewards along the upper margin during the medial excursion of the vocal fold[10,30-32] (Figure 1).

The sharpness and roundedness as observed from the shape of the lateral peaks are resulting from the vertical phase differences between the lower and upper margins of the vibrating vocal folds[30,33] (Figure 2). Looking from above the vocal folds, the boundary between the glottis and the vocal fold is created by the most medial part of the vocal fold. Due to the vertical phase differences, during the opening phase, this boundary is formed by the position of the upper margin of the vocal fold, whereas during the closing phase the boundary is normally formed by the position of the lower margin of the vocal fold. At the point of transition from opening to the closing phase, the glottal edge shifts from the upper to the lower margin (Figure 2A). When the vertical phase differences are large, the shift from upper to lower vocal fold margin is abrupt. In the kymogram, this sudden transition results in a sharp lateral peak within the oscillating vocal fold contour. In smaller vertical phase differences this transition happens rather gradually, causing the lateral peak to be rounded (Figure 2B).

The shape of the lateral peak has been found to be a clinically useful parameter revealing the vocal fold vibration characteristics that are not easily observable in non-kymographic imaging methods.[34] It has gained attention due to its diagnostic importance in assessing various voice disorders such as mucosal inflammations, scarring or tumors related to increased mucosal stiffness.[10,30,31,34-37] Increased vertical thickness of the vocal folds and increased pliability of the mucosa likely lead to larger vertical phase differences producing sharper lateral peaks in kymography. In contrast, increased stiffness of the mucosa is expected to reduce vertical phase differences, thus producing more rounded lateral peak shapes.[10,30,32] The magnitude of the vertical phase differences and the related sharpness of lateral peaks can also reveal on vocal fold vibratory behavior in different voice tokens, such as vocal registers.[33,38]

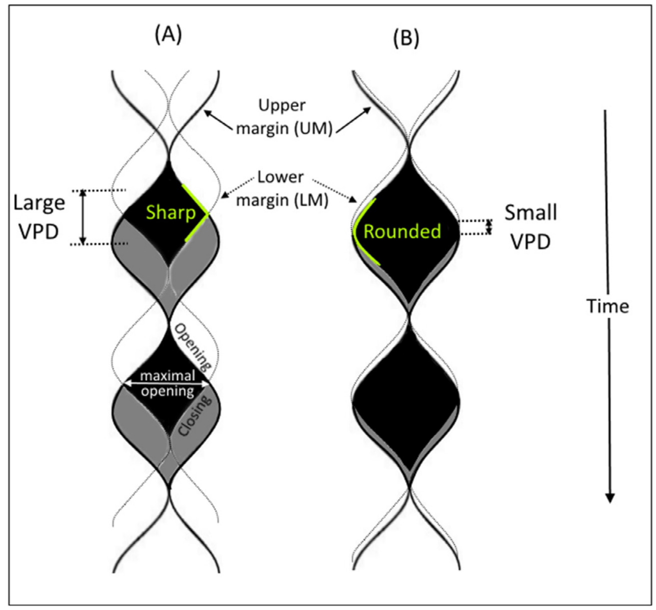


FIGURE 2. Formation of sharp (A) and rounded (B) lateral peaks in the kymogram. Movements of the lower and upper margins of the vocal folds are indicated by thin-dotted and thick-solid curves, respectively. The vibratory displacement of the lower margin precedes that of the upper margin, thus creating a vertical phase difference between their respecive motions. During the opening phase, the motion of the lower margin is invisible—this is indicated by the thin-dotted line; it becomes only visible during the closing phase. A sharp lateral peak (A) is seen when the vertical phase difference is large and a rounded lateral peak (B) is seen when the vertical phase difference is small (indicated in green). LM, lower margin; UM, upper margin; VPD, vertical phase difference. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

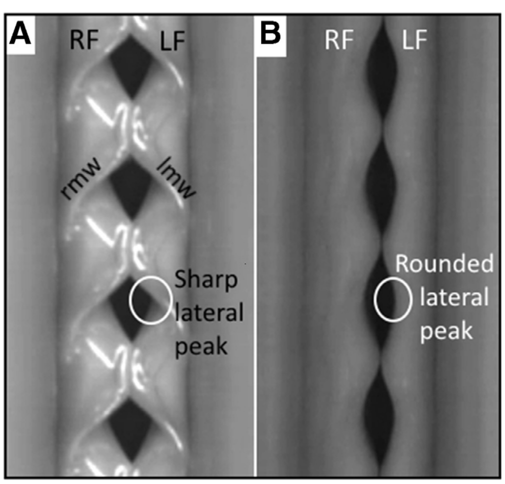


FIGURE 1. Videokymographic images (four vibratory cycles each) showing (A) sharp lateral peaks (encircled) and laterally running mucosal waves (rmw, lmw) on the right and left vocal fold, respectively; (B) rounded lateral peaks (encircled) with no mucosal waves. RF, LF – right and left vocal fold. Total time displayed in the kymograms: 17.6 ms (time direction from top to bottom).

Efforts have been made to assess vertical phase differences and laterally traveling mucosal waves using image analysis methods. Shaw and Deliyski (2008)[39] used mucosal wave playback and qualitatively assessed the variations in mucosal wave magnitude and symmetry. Voigt et al (2010)[23] managed to detect the laterally traveling mucosal waves in high-speed endoscopic videos using automated image analysis techniques. Lately, Andrade-Miranda et al (2017)[40] used the optical flow method to detect mucosal wave propagation from high-speed endoscopic videos.

Chen, Woo, and Murry[41-43] applied spectral analysis to vocal fold waveforms obtained from digital kymograms and reported its usefulness in quantifying the waveforms and their changes due to different vocal tokens, pathologies, and surgical interventions. In principle, the spectral features can be expected to reflect the sharpness of the lateral peaks through

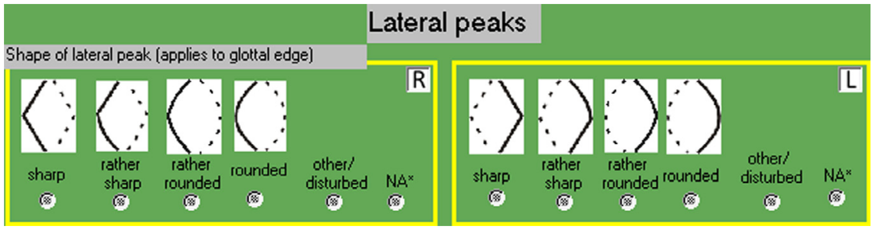


FIGURE 3. The form for visual evaluation of the VKG images with the descriptive pictograms representing varied degrees of lateral peak sharpness in the right (R) and left (L) vocal folds.

increased energy in upper harmonics, but such spectral changes can occur also due to, eg, the occurrence of closed phase; thus the spectral analysis of kymographic waveform makes it difficult to clearly distinguish the sharpness of the lateral peaks from other factors.

According to our knowledge, two methods have tried to quantify the shape of the lateral peak from kymograms so far.[44,45] Jiang et al estimated the shape of the peak indirectly by quantifying the vertical phase difference from kymographic images using a sinusoidal model approximation.[44,46-51] While this method is mathematically elegant, it becomes troublesome and difficult to interpret when the vocal fold motion becomes rather complex. The second method by Yamauchi et al (2015) quantified the peak sharpness from digital kymograms by the “lateral peak index”, defined as an angle formed by two lines between the start of open phase and lateral peak, and between the lateral peak and the end of open phase.[45] However, this index disregards the changes of curvature of the vocal fold waveform that influence the peak sharpness. Its value is additionally influenced also by the closed quotient and the vibratory amplitude, thus making it also sensitive to other factors than vertical phase differences.

Therefore, there is a need to search for other parameters that could help to improve the reliability of visual evaluation of clinical kymographic images of the vocal folds. Due to limited inter- and intrarater reliability, the approach of subjective rating limits the comparability of quantitative parameters on a large set of data. It further prevents the acquisition of reliable standard reference values for clinicians whose treatment decisions are dependent on assessment of such parameters. In contrast, objectification helps to find the accuracy of visual ratings.[52] The purpose of this study was therefore to investigate parameters which could quantify the lateral peak sharpness seen in the kymographic images and could easily be measured automatically from the detected contours of the vibrating vocal folds.

The work was done in the following steps: (1) A set of clinically obtained videokymographic images was evaluated visually to obtain ratings of the lateral peak sharpness. (2) The same images were subjected to automatic image analysis, in order to detect and compute the contours of the vibrating vocal folds as waveforms. (3) The resulting waveforms were quantified in order to obtain numerous parameters expected to reflect the lateral peak sharpness. (4) The obtained values of the parameters were compared to the visual ratings from step (1), in order to determine the parameters that show the best correlation with the visual ratings.

METHODS

Dataset

The dataset used in this work consisted of 45 videokymographic (VKG) images retrospectively selected from clinical records of patients examined for voice complaints at the Voice and Hearing Centre, Medical Healthcom, Ltd, Prague. The VKG recordings were obtained with the second generation VKG camera (Kymocam, CYMO, b.v. Groningen, the Netherlands, image rate 7200 lines/s), which was connected to a laryngoscope (Xion Medical, Germany, 10 mm diameter, 90° angle) using a C-mount objective adapter (R. Wolf, Germany, type 85261.272, 27 mm focal length). The larynx was illuminated by a 300 W endoscopic xenon light source (type FX 300 A, Fentex Medical, Germany). The VKG recordings were stored digitally by means of an EndoSTROB video capturing unit (Xion Medical, Germany). The images were extracted from the video records using the recently developed VKG Analyzer software.[53] The images were selected so that they demonstrated varied degrees of sharpness of lateral peaks.

Visual rating

Three raters independently evaluated the sharpness of the lateral peaks from the VKG images using a visual form (Figure 3).[54] The raters used the pictogram descriptions of the sharpness features as a reference for evaluation. The rating was done on a four points rating scale (1-sharp; 2-rather sharp; 3-rather rounded; 4-rounded) for left and right vocal folds separately, thus making a total of 90 ratings per rater from 45 images.

In order to assess the intra-rater reliability, each rater performed the evaluation twice, with a pause of 7-10 days in between. During the second evaluation, the order of the images was changed to minimize the memory effect. The ratings from the two evaluations, for the three raters, were consolidated, and an average (visual average – VA) was obtained. A common consensus (visual consensus – VC)

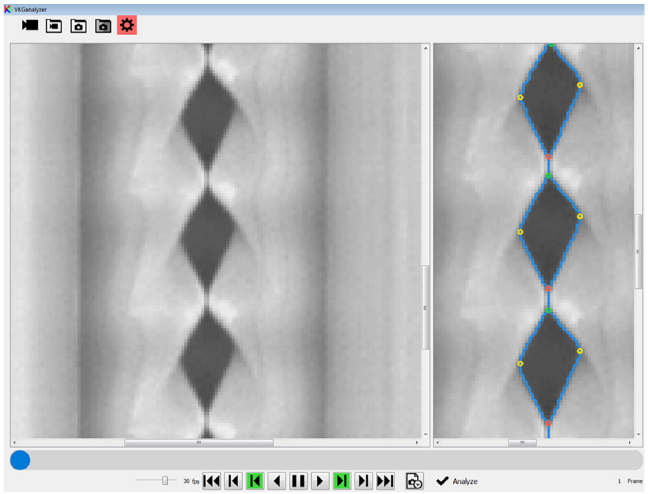


FIGURE 4. The screenshot of the VKG analyzer software showing the VKG image on the left and the detected glottal edge contours on the right.

was also arrived through the discussion among the three raters afterwards.

Image analysis

The recently developed VKG analyzer software[53] was used to detect and extract the contours defining the glottal edge boundary of both the left and right vocal folds (Figure 4). The image brightness and contrast were manually adjusted to improve the accuracy of the edge detection whenever required. The contours extracted from each of the VKG images were saved in a text file as a set of data defining the glottal edges of the left and right vocal folds, along with their respective time instances. A custom MATLAB script was then used to process the vocal fold contours and to obtain parameters capturing lateral peak sharpness, which could be included in the VKG analyzer software in future versions.

Quantification of lateral peak sharpness

Two kinds of parameters were defined for their simplicity in quantifying the vocal fold waveforms and their expected capability of reflecting the sharpness of the lateral peaks: the *Open Time Percentage Quotients* (OTQ) and *Plateau Quotients* (PQ).

The *Open Time Percentage Quotients* (OTQ_R) were inspired by the OT50 parameter published by Woo (1996),[55] who investigated the time for which the glottal area waveform exceeded 50% of the amplitude. Here, we defined the OTQ_R parameter as the proportion of time during which the vocal fold displacement exceeds a chosen percentage (R) of the vibration amplitude within a period (Figure 5):

$$OTQ_R = \frac{D_R}{T}$$

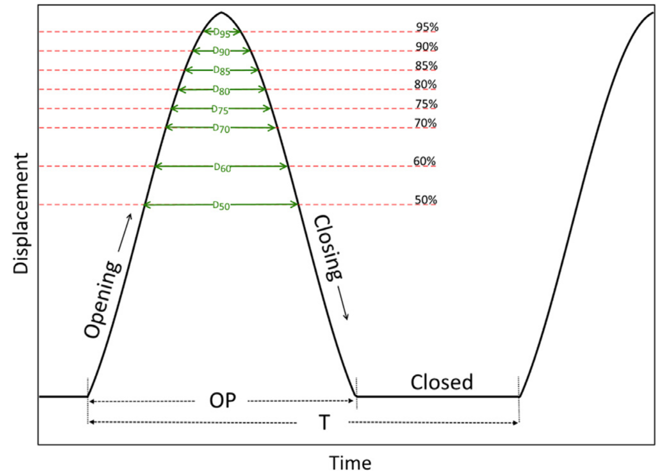


FIGURE 5. Parameterization of the vocal fold waveform for obtaining the Open Time Percentage Quotients (OTQ_R) and the Plateau Quotients (PQ_R) as indicators for peak sharpness. OP is the open phase, T is the period, and D_R is the duration of the phase during which the waveform exceeds a specified R percentage of the amplitude. The R percentages are indicated by the dashed red lines.

where D_R is the duration of the phase where the lateral displacement is greater than R% of the vibration amplitude and T is the period of the vocal fold vibratory cycle. The vibration amplitude was determined as the difference between the most lateral and most medial position of the vocal fold during the open phase.

The *Plateau Quotients* (PQ_R) used here were inspired by the work of Mehta et al,[56] who investigated the proportion of open phase for which the glottal area was larger than 95% of its maximum. Here, we defined PQ_R as the proportion of time during which the vocal fold displacement

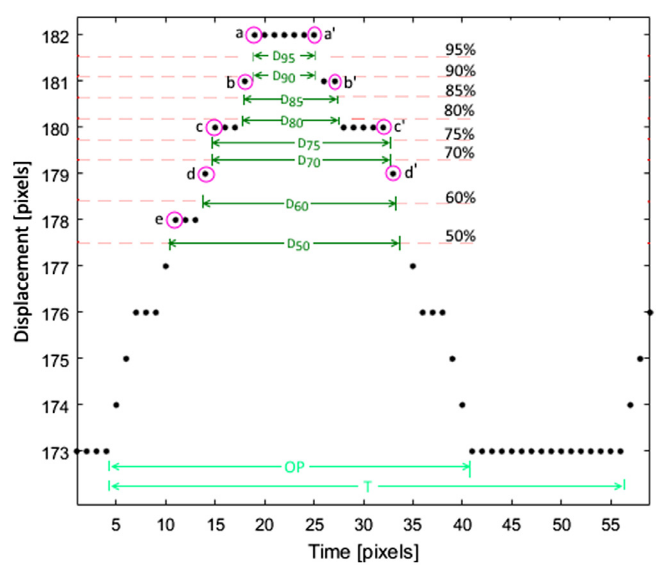


FIGURE 6. Implementation of the parameterization of the waveform illustrating the procedure followed to calculate the D_R durations from the discrete samples.

exceeds R% of vibration amplitude within the open phase (Figure 5):

$$PQ_R = \frac{D_R}{OP}$$

where OP is the duration of the open phase.

When implementing the automatic analysis procedure, it was necessary to deal with the fact that the waveforms were not continuous, but consisted of samples of limited temporal and spatial resolution. While the contour samples are defined by integer pixel coordinates, the R% levels usually correspond to noninteger subpixel coordinates. An example of the procedure adapted to calculate the OTQ and PQ parameters from the discrete samples is shown in Figure 6. When digitized, the discrete contour data points were located at specific pixels with coordinates defined by integer numbers. Therefore, sometimes the same pixel coordinates pertained to multiple consecutive time points (see Figure 6). In order to measure the time intervals at which the vocal fold displacement exceeds the criterion level R%, the first and last samples with the values above the R% criterion level in the opening and closing phases, respectively, were selected (marked by circles and indicated as a, b, c, d, e in the opening phase, and a', b', c', d' in the closing phase in Figure 6). Thus, for the R% levels at 95%, 90%, 85%, 80%, 75%, 70%, 60% and 50%, the intervals between $a-a'$, $a-a'$, $b-b'$, $b-b'$, $c-c'$, $c-c'$, $d-d'$ and $e-d'$, respectively, were considered to calculate the D_R durations in the example shown in Figure 6.

Statistical analysis

Statistical analysis was performed using the SPSS (version 24) software. Spearman's rank correlation coefficient was computed to determine the inter- and intrarater reliability of the visual ratings. The intrarater reliability was also tested with Cronbach's Alpha value. To estimate the correlation between the objective measures and the visual ratings, Spearman's rank correlation coefficient was again used.

RESULTS

Visual rating

Results from the repeated visual evaluations of the lateral peak sharpness in VKG images by the three raters were compared to find the intrarater and inter-rater reliability. The intrarater comparisons between the two repeated evaluations resulted in the Cronbach's Alpha values around 0.92 for all three raters, indicating excellent reliability of the raters. The intrarater Spearman's rank correlation coefficients for the individual raters varied between 0.84 and 0.85 ($P < 0.001$, N = 90) indicating very strong and significant correlations between the repeated evaluations.

The inter-rater comparisons showed Spearman's rank correlation coefficients in the range of 0.67 to 0.82, with a mean value of 0.73. These coefficients indicated strong and significant correlations ($P < 0.001$, N = 90) between the evaluations of the different raters, but also hinted at some discrepancies among the raters. Therefore, a consensus among the raters was established by mutual discussions.

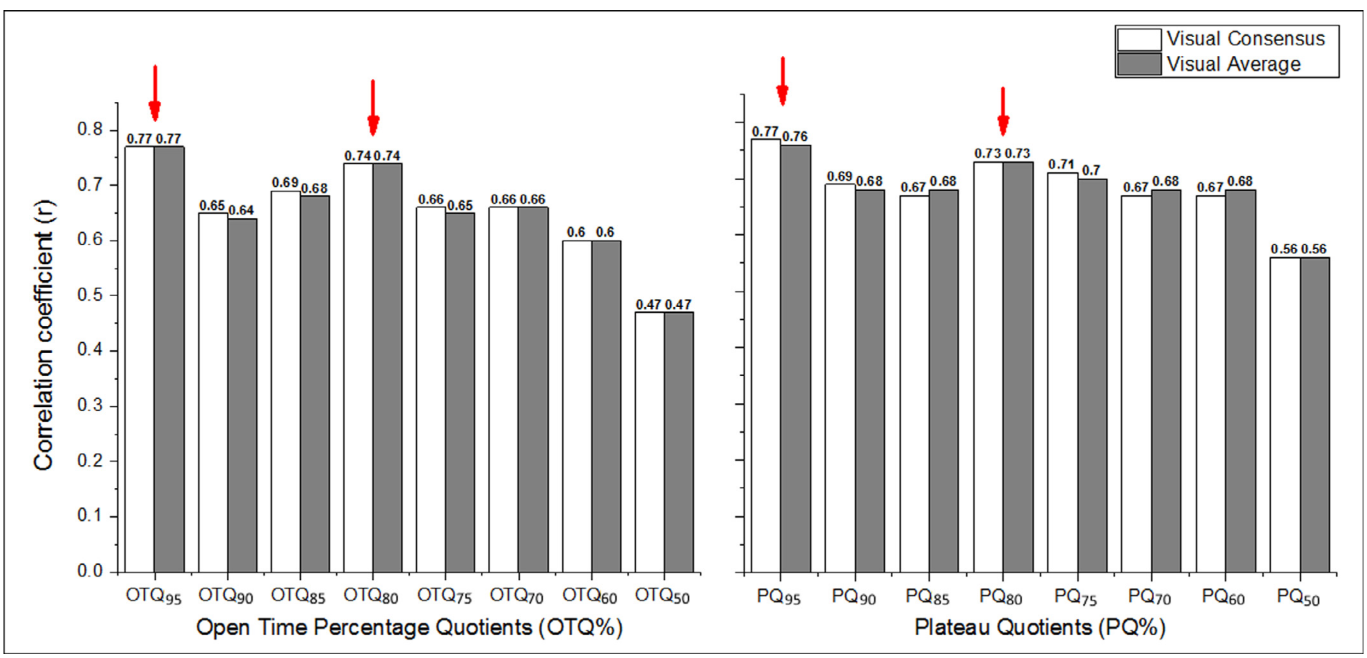


FIGURE 7. Spearman's rank correlation coefficients indicating the agreement between the visual ratings and the measured parameters OTQ and PQ. The highest correlation coefficients are indicated by red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The visual consensus versus visual average comparison revealed very strong Spearman's rank correlation ($r = 0.99$, $P < 0.001$). Furthermore, both the visual consensus and visual average values very strongly correlated with the values of all three raters ($r = 0.81$-0.91, $P < 0.001$) in both evaluations. Therefore, the visual consensus and visual average values were deemed appropriate for further analysis of the correlations between the visual and automatic image analysis.

Correlation between visual ratings and the analyzed parameters

The correlations between the different OTQ and PQ parameters with the visual consensus and visual average ratings are shown in Figure 7. All correlations had a significance level of $P < 0.001$, indicating that all parameters were well related to the visual ratings. Highest correlations were found for the parameters measured at 95% amplitude (OTQ_{95}, PQ_{95}) and at 80% amplitude (OTQ_{80}, PQ_{80}). In Figure 7, these are indicated by arrows. There were minimal differences between the OTQ and PQ parameters measured at the same percentage. Also, there were minimal differences between the visual average and visual consensus. Lowest correlations were found for the parameters measured at 50% amplitude (OTQ_{50}, PQ_{50}).

The relationships between the values of the four best correlating parameters OTQ_{95}, OTQ_{80}, PQ_{95}, and PQ_{80}, and the visual ratings are revealed in Figure 8. As expected, all these quotients clearly increase their values when the peak shape changes from sharp to rounded. There is, however, some spread of the measured data around the best fit line, which indicates that some discrepancies exist between the visual and automatic evaluations. The Spearman's rank correlation values between

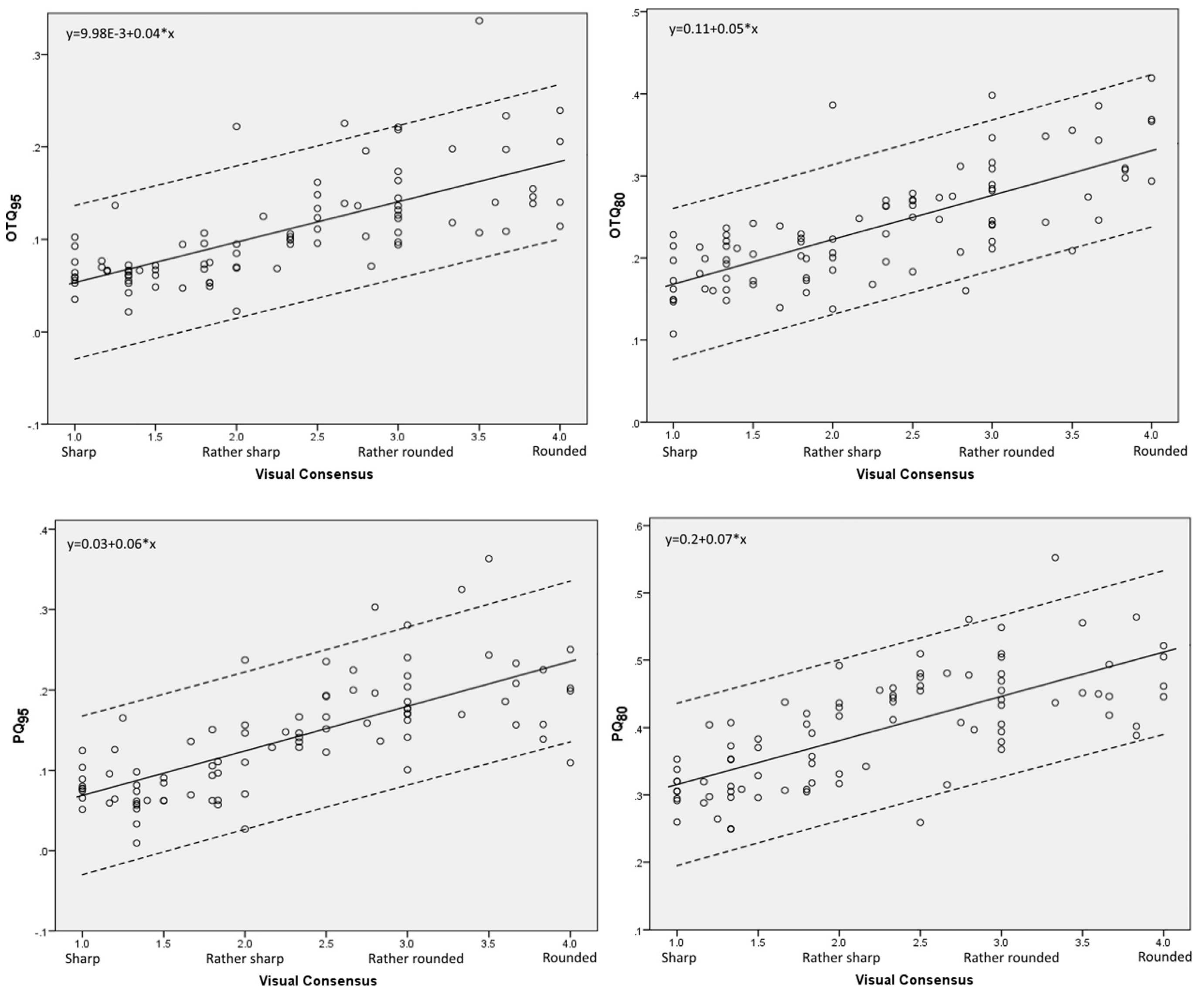


FIGURE 8. The relationship between the measured values and the visual ratings for the four parameters with the highest correlations – OTQ_{95}, OTQ_{80}, PQ_{95}, and PQ_{80}. The lines indicate the best fit linear relationship (solid) and 95% confidence intervals (dashed).

these analyzed quotients and the visual ratings (0.73-0.77, as shown in Figure 7) were comparable to those found between different raters (0.67-0.82) indicating that the discrepancies in the automatic-to-visual comparisons are similar to those found in inter-rater comparisons.

DISCUSSION

Sharpness of lateral peaks has been recognized previously as a useful visual feature that can indicate pliability and health of the vocal fold mucosa.[30,45] In a recent study, the lateral peak sharpness has been identified as one of the most helpful visual features for clinical evaluation of voice disorders using videokymography.[34] The peak sharpness is directly related to vertical phase differences between the motions of the upper and lower margin of the vocal folds and results from projection of the vocal fold motion into the laryngoscopic view from above of the vocal folds.[10,30,57] Biomechanically, stiffening of the mucosa leads to increased mucosal wave speed[6] and decreased vertical phase differences, causing the peak to become more rounded.[30,32,35] Apart from physiological factors related, eg, to pitch increase and voice registration, stiffening of the mucosa is considered to be a direct result of pathological processes on the vocal folds. Therefore, evaluation of peak sharpness can help clinicians to better diagnose the health of the vocal fold mucosa, particularly in phonations produced at comfortable pitch in modal/chest register where the mucosa is expected to be pliable.

Visual evaluation, however, is subjective and differences among evaluations of different raters can be expected. This can be spotted also in our results: while the intraindividual Spearman's rank correlations were very strong ($r = 0.84$-0.85), the inter-rater Spearman's rank correlations were lower ($r = $ -0.67-0.82) indicating more disagreements between the visual evaluations of different raters than between repeated evaluations of the same rater.

This study searched for objective parameters that are related to the visual ratings of peak sharpness in kymograms and can be used as "peak sharpness indicators". For this purpose, the OTQ and PQ were defined by relating the durations of different phases of the vibratory cycle to each other, applying the same concept as used for the well-established traditional parameters such as the Closed Quotient (CQ), Open Quotient (OQ) or Speed Quotient (SQ).[58-60] As such, these parameters are relatively simple to measure. As far as their interpretation is concerned, smaller OTQ and PQ values correspond to sharper lateral peaks of the vocal fold waveform detected in the kymogram (recall Figure 8).

The OTQ and PQ parameters measured from the time intervals at different percentages of vibratory amplitude were compared to the visual ratings of the peak sharpness in order to evaluate the congruence of these two approaches. The OTQ and PQ parameters showed very similar correlations to the visual ratings which indicate that they quantified the visual impressions similarly. The best correlations with the visual evaluations were found for the OTQ and PQ parameters measured at 95% and 80% of the amplitude, the worst correlations appeared at 50% of the amplitude. Since the peak corresponds to 100% of the amplitude, it appears logical that the best correlations for peak sharpness should be obtained for the measurements made as closely to the peak as possible − this explains the finding of the worst correlations at 50% and best correlations at 95% of the amplitude (recall Figure 7). However, the correlations at 90% and 85% of the amplitude were worse than those at 80%. This seemingly contradictory finding could be attributed to the contour artifacts due to the limited pixel and temporal resolution (compare the ideal waveform in Figure 5 with the real detected waveform in Figure 6). The clinical videokymographic images analyzed here showed the average vocal fold vibratory amplitudes around 8 pixels (range 5-15 pixels). A change of 1 pixel, in this case, corresponds to the spatial resolution of 12.5% of the amplitude (range 7-20%). This means that it is hardly possible to reliably distinguish levels that are close together, such as those at 85%, 90%, and 95% of the amplitude.

Preliminary investigations using synthetic kymograms generated by a kinematic model of the vocal folds[61] with known vertical phase differences (not included here for brevity reasons) showed that the limited spatial and temporal resolution of the kymographic images can influence the accuracy of the results, particularly of those quotients measured at the proximity of the peak, and these artifacts need to be taken into account. Thus the measurements at 80% amplitude could potentially be used as a compromise to reduce the influence of the possible waveform artifacts, but still reflect the peak sharpness and vertical phase differences reasonably well. The waveform artifacts present a general limitation which is inherent in the laryngeal kymographic techniques. Increased spatial resolution of the kymographic images is desirable for improving the quantification accuracy of the vocal fold vibratory patterns in future.

In principle, the OTQ and PQ parameters can be implemented also for analyzing the glottal area waveforms (GAWs) obtained from full high-speed endoscopic videos, as done by Mehta et al (2011).[56] GAWs offer better pixel resolution than kymography due to the fact that the glottal area is distributed over multiple image lines and thus over considerably more pixels. In this respect, GAWs may possibly offer better accuracy than kymographic waveforms in measuring the OTQ and PQ parameters as defined here. However, a more detailed study is needed to elucidate these factors and to better understand the influence of limited spatial and temporal resolution on the accuracy of these parameters.

The detailed comparisons between the visual ratings and the OTQ and PQ parameters shown in Figure 8 reveal that the relationship is not perfect and some discrepancies exist here. Besides of the influence of the limited spatial and temporal resolution of the images (7200 kymographic lines per second with 720 pixels per line used here), these discrepancies could possibly be also due to contour detection artifacts resulting from the image analysis procedure. Furthermore, it is known that the visual perception process is rather complex and visual judgments of the peak shape may also be influenced by, eg, the grayscale shadings which are not captured in the contours. All these factors may contribute to the differences between the automatic analysis and the visual ratings. Nevertheless, the Spearman's rank correlations between the visual ratings and the OTQ and PQ parameters measured at 80% and 95% amplitude ($r = 0.73$-0.77, recall Figure 7) are similar to those found between different raters. Therefore the reliability of the parameters, although not perfect, is considered acceptable here.

While the shape of the lateral peak appears as a useful clinical feature, ultimately it should be related to the vertical phase differences. These differences cannot be exactly measured laryngoscopically *in vivo*. Therefore, we were not able to establish their direct relationship with the defined parameters, which poses another potential limitation of this study. However, this relationship may be derived and investigated using synthetic kymograms obtained from a mathematical model of the vocal folds with known vertical phase differences[61], which is planned to be addressed in a future study.

CONCLUSION

The PQ_{95}, PQ_{80}, OTQ_{95} and OTQ_{80} parameters stood out as the possible candidates for capturing the sharpness of the lateral peaks. The reliability of these parameters appears comparable to the inter-individual reliability of visual ratings. The results provide basic insights into developing the computer algorithms to automatically quantify the sharpness of lateral peaks from the VKG images.

REFERENCES

1. Hirano M. *Clinical Examination of Voice*. Wien, Austria: Springer-Verlag; 1981.
2. Titze IR. The physics of small−amplitude oscillation of the vocal folds. *J Acoust Soc Am*. 1988;83:1536–1552.
3. McGowan R. An analogy between the mucosal waves of the vocal folds and wind waves on water. *Haskins Lab Status Rep Speech Res*. 1990;101:243–249.
4. Yumoto E, Kurokawa H, Okamura H. Vocal fold vibration of the canine larynx: observation from an infraglottic view. *J Voice*. 1991;5:299–303.
5. Titze IR, Jiang JJ, Hsiao T-Y. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. *Ann Otol Rhinol Laryngol*. 1993;102:58–63.
6. Berke GS, Gerratt BR. Laryngeal biomechanics: an overview of mucosal wave mechanics. *J Voice*. 1993;7:123–128.
7. Boessenecker A, Berry DA, Lohscheller J, et al. Mucosal wave properties of a human vocal fold. *Acta Acust united Ac*. 2007;93:815–823.
8. Krausert CR, Olszewski AE, Taylor LN, et al. Mucosal wave measurement and visualization techniques. *J Voice*. 2011;25:395–405.
9. Hirano M, Bless DM. *Videostroboscopic Examination of the Larynx*. San Diego, California: Singular Publishing Group; 1993.
10. Švec JG, Šram F, Schutte HK. Videokymography. In: Fried M, Ferlito A, eds. 3 ed*The Larynx*. Vol 1, San Diego, CA: Plural Publishing; 2009:253–271.
11. Bless DM, Hirano M, Feder RJ. Videostroboscopic evaluation of the larynx. *Ear Nose Throat J*. 1987;66:289–296.
12. Hiroto I. Vibration of vocal cords: an ultra high-speed cinematographic study (film). Kurume, Japan: Department of otolaryngology, Kurume University; 1968.
13. Berry DA, Montequin DW, Tayama N. High-speed digital imaging of the medial surface of the vocal folds. *J Acoust Soc Am*. 2001;110:2539–2547.
14. Döllinger M, Berry DA, Kniesburges S. Dynamic vocal fold parameters with changing adduction in ex-vivo hemilarynx experiments. *J Acoust Soc Am*. 2016;139:2372–2385.
15. Herbst CT, Hampala V, Garcia M, et al. Hemi-laryngeal setup for studying vocal fold vibration in three dimensions. *J Vis Exp*. 2017; 129:e55303. http://dx.doi.org/10.3791/55303.
16. Jing B, Ge Z, Wu L, et al. Visualizing the mechanical wave of vocal fold tissue during phonation using electroglottogram-triggered ultrasonography. *J Acoust Soc Am*. 2018;143:EL425–EL429.
17. Ishizaka K, Flanagan J. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst Tech J*. 1972;51:1233–1268.
18. Titze IR. Comments on the myoelastic-aerodynamic theory of phonation. *J Speech Hear Reas*. 1980;23:495–510.
19. Titze IR. *Principles of Voice Production (Second Printing)*. Iowa City, IA: National Center for Voice and Speech; 2000.
20. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur Arch Otorhinolaryngol*. 2001;258:77–82.
21. Poburka BJ. A new stroboscopy rating form. *J Voice*. 1999;13: 403–413.
22. Deliyski DD, Petrushev PP, Bonilha HS, et al. Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. *Folia Phoniatr Logop*. 2008;60:33–44.
23. Voigt D, Döllinger M, Eysholdt U, et al. Objective detection and quantification of mucosal wave propagation. *J Acoust Soc Am*. 2010;128: EL347–EL353.
24. Kaneko M, Shiromoto O, Fujiu-Kurachi M, et al. Optimal duration for voice rest after vocal fold surgery: randomized controlled clinical study. *J Voice*. 2017;31:97–103.
25. Poburka BJ, Patel RR, Bless DM. Voice-vibratory assessment with laryngeal imaging (VALI) form: reliability of rating stroboscopy and high-speed videoendoscopy. *J Voice*. 2017;31:513e1–513.e14.
26. El-Demerdash A, Fawaz SA, Sabri SM, et al. Sensitivity and specificity of stroboscopy in preoperative differentiation of dysplasia from early invasive glottic carcinoma. *Eur Arch Otorhinolaryngol*. 2015;272:1189–1193.
27. Zacharias SRC, Deliyski DD, Gerlach TT. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *J Voice*. 2018;32:216–220.
28. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended minimum protocols for instrumental assessment of voice: American Speech-Language Hearing Association Committee on Instrumental Voice assessment protocols. *Am J Speech Lang Pathol*. 2018;27:887–905.
29. Švec JG, Schutte HK. Kymographic imaging of laryngeal vibrations. *Curr Opin Otolaryngol Head Neck Surg*. 2012;20:458–465.
30. Švec JG, Šram F, Schutte HK. Videokymography in voice disorders: what to look for. *Ann Otol Rhinol Laryngol*. 2007;116:172–180.
31. Švec JG, Frič M, Šram F, Schutte HK. Mucosal waves on the vocal folds: conceptualization based on videokymography. *Fifth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*. Firenze, Italy: Firenze University Press; 2007:171–172.
32. Švec JG, Šram F. Videokymographic examination of voice. In: Ma EPM, Yiu EML, eds. *Handbook of Voice Assessments*. San Diego, CA: Plural Publishing; 2011:129–146.
33. Sundberg J, Högset C. Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. *Logoped Phoniatr Vocol*. 2001;26:26–36.

34. Phadke KV, Vydrová J, Domagalská R, et al. Evaluation of clinical value of videokymography for diagnosis and treatment of voice disorders. *Eur Arch Otorhinolaryngol*. 2017;274:3941–3949.
35. Vydrová J, Švec JG, Šram F. Videokymography (VKG) in laryngologic practice. *J Macrotrends Health Med*. 2015;3:87–95.
36. Yamauchi A, Yokonishi H, Imagawa H, et al. Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging. *J Voice*. 2016;30:205–214.
37. Yamauchi A, Yokonishi H, Imagawa H, et al. Visualization and estimation of vibratory disturbance in vocal fold scar using high-speed digital imaging. *J Voice*. 2016;30:493–500.
38. Švec JG, Sundberg J, Hertegard S. Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography. *J Acoust Soc Am*. 2008;123:347–353.
39. Shaw HS, Deliyski DD. Mucosal wave: a normophonic study across visualization techniques. *J Voice*. 2008;22:23–33.
40. Andrade-Miranda G, Bernardoni NH, Godino-Llorente JI. Synthesizing the motion of the vocal folds using optical flow based techniques. *Biomed Signal Process Control*. 2017;34:25–35.
41. Chen W, Woo P, Murry T. Spectral analysis of digital kymography in normal adult vocal fold vibration. *J Voice*. 2014;28:356–361.
42. Chen W, Woo P, Murry T. Vocal fold vibratory changes following surgical intervention. *J Voice*. 2016;30:224–227.
43. Chen W, Woo P, Murry T. Vocal fold vibration following surgical intervention in three vocal pathologies: a preliminary study. *J Voice*. 2017;31:610–614.
44. Jiang JJ, Chang CIB, Raviv JR, et al. Quantitative study of mucosal wave via videokymography in canine larynges. *Laryngoscope*. 2000;110:1567–1573.
45. Yamauchi A, Yokonishi H, Imagawa H, et al. Quantitative analysis of digital videokymography: a preliminary study on age- and gender-related difference of vocal fold vibration in normal speakers. *J Voice*. 2015;29:109–119.
46. Jiang JJ, Zhang Y, Kelly MP, et al. An automatic method to quantify mucosal waves via videokymography. *Laryngoscope*. 2008;118: 1504–1510.
47. Chodara AM, Krausert CR, Jiang JJ. Kymographic characterization of vibration in human vocal folds with nodules and polyps. *Laryngoscope*. 2012;122:58–65.
48. Krausert CR, Ying D, Zhang Y, et al. Quantitative study of vibrational symmetry of injured vocal folds via digital kymography in excised canine larynges. *J Speech Lang Hear Res*. 2011;54:1022–1038.
49. Li L, Zhang Y, Maytag AL, et al. Quantitative study for the surface dehydration of vocal folds based on high-speed imaging. *J Voice*. 2015;29:403–409.
50. Regner MF, Robitaille MJ, Jiang JJ. Interspecies comparison of mucosal wave properties using high-speed digital imaging. *Laryngoscope*. 2010;120:1188–1194.
51. Zhang Y, Huang N, Calawerts W, et al. Quantifying the subharmonic mucosal wave in excised larynges via digital kymography. *J Voice*. 2017;31:123.e7–123.e13.
52. Bonilha HS, Deliyski DD, Gerlach TT. Phase asymmetries in normophonic speakers: visual judgments and objective findings. *Am J Speech Lang Pathol*. 2008;17:367–376.
53. Novozamsky A, Sedlar J, Zita A, et al. Image analysis of videokymographic data. *2015 IEEE International Conference on Image Processing (ICIP)*, 2015,78−82.
54. Švec JG, Švecová H, Herbst C, et al. Evaluation protocol for videokymographic images. (Ms Access software application). Groningen, the Netherlands: Groningen Voice Research Lab, University of Groningen. 2007.
55. Woo P. Quantification of videostrobolaryngoscopic findings−measurements of the normal glottal cycle. *Laryngoscope*. 1996;106:1–27.
56. Mehta DD, Zañartu M, Quatieri TF, et al. Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy. *J Speech Lang Hear Res*. 2011;130:3999–4009.
57. Hiroto I. The mechanism of phonation; its pathophysiological aspects. *Nippon Jibiinkoka Gakkai Kaiho*. 1966;69:2097–2106.
58. Timcke R, von Leden H, Moore P. Laryngeal vibrations - measurements of the glottic wave .I. The normal vibratory cycle. *AMA Arch Otolaryngol*. 1958;68:1–19.
59. Qiu Q, Schutte H, Gu L, et al. An automatic method to quantify the vibration properties of human vocal folds via videokymography. *Folia Phoniatr Logop*. 2003;55:128–136.
60. Lohscheller J, Švec JG, Döllinger M. Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects. *Logoped Phoniatr Vocol*. 2013;38:182–192.
61. Subbaraj PK, Švec JG. Kinematic model for simulating mucosal wave phenomena on vocal folds. In: Manfredi C, ed. *MAVEBA 2017: Models and Analysis of Vocal Emissions for Biomedical Applications. 10th International Workshop*. Firenze: Firenze University Press; 2017:115–118.

Supplement C: Manuscript III

Cepstral and Perceptual Investigations in Female Teachers with Functionally Healthy Voice

Authors and their contribution to the study in percentages:

K.V. Phadke	**35 %**
A.M. Laukkanen	20 %
I. Ilomäki	10 %
E. Kankare	10 %
A. Geneid	5 %
J.G. Švec	20 %

Journal name and Impact factor (IF):

Journal of Voice, IF (2016): 1.381

Authors' contribution:

K.V. Phadke: Analyzed the results, applied statistics, wrote the manuscript

A.M. Laukkanen: Initiated the study in Finland, supervised the collection of the teachers' data, analyzed the voice recordings, and revised the manuscript.

I. Ilomäki: Collected the data (recorded the voices of teachers in schools) and performed perceptual ratings of the voice of teachers. Approved the final version of the manuscript.

E. Kankare: Collected the data (recorded the voices of teachers in schools) and performed perceptual ratings of the voice of teachers. Approved the final version of the manuscript.

A. Geneid: Organized and supervised the stay of K.V. Phadke in Finland and her participation in the study. Corrected the final version of the manuscript.

J.G. Švec: Critically revised the study and designed the methodology of analysis of the voice recordings, analyzed the results and verified the accuracy of the data, supervised the writing and critically revised the successive versions of the manuscript.

Cepstral and Perceptual Investigations in Female Teachers With Functionally Healthy Voice

*Ketaki Vasant Phadke, †Anne-Maria Laukkanen, †Irma Ilomäki, ‡Elina Kankare, §Ahmed Geneid, and *Jan G Švec, *Olomouc, Czech Republic, and †‡Tampere, and §Helsinki, Finland

Abstract: Purpose. The present study aimed at measuring the smoothed and non-smoothed cepstral peak prominence (CPPS and CPP) in teachers who considered themselves to have normal voice but some of them had laryngeal pathology. The changes of CPP, CPPS, sound pressure level (SPL) and perceptual ratings with different voice tasks were investigated and the influence of vocal pathology on these measures was studied.

Method. Eighty-four Finnish female primary school teachers volunteered as participants. Laryngoscopically, 52.4% of these had laryngeal changes (39.3% mild, 13.1% disordered). Sound recordings were made for phonations of comfortable sustained vowel, comfortable speech, and speech produced at increased loudness level as used during teaching. CPP, CPPS and SPL values were extracted using *Praat* software for all three voice samples. Sound samples were also perceptually evaluated by five voice experts for overall voice quality (10 point scale from poor to excellent) and vocal firmness (10 point scale from breathy to pressed, with normal in the middle).

Results. The CPP, CPPS and SPL values were significantly higher for vowels than for comfortable speech and for loud speech compared to comfortable speech ($P < 0.001$). Significant correlations were found between SPL and cepstral measures. The loud speech was perceived to be firmer and have a better voice quality than comfortable speech. No significant relationships of the laryngeal pathology status with cepstral values, perceptual ratings, or voice SPLs were found ($P > 0.05$).

Conclusion. Neither the acoustic measures (CPP, CPPS, and SPL) nor the perceptual evaluations could clearly distinguish teachers with laryngeal changes from laryngeally healthy teachers. Considering no vocal complaints of the subjects, the data could be considered representative of teachers with functionally healthy voice.

Key Words: Teachers' voice−Voice SPL−CPP−CPPS−Perceptual evaluation−Laryngeal pathologies.

INTRODUCTION

Cepstral peak prominence (CPP) and the smoothed cepstral peak prominence (CPPS)[1,2] are considered to be rather robust acoustic measures of overall severity of dysphonia.[3,4] CPP is a measure of the relative cepstral peak amplitude (in decibels) of the voice signal.[1,2] It is obtained by finding out the difference between the maximum cepstral peak value occurring within the boundaries of the expected phonational quefrencies and the corresponding value on the regression line fitted on the cepstrum. CPP was originally developed to analyze sustained vowels and measure the degree of harmonic organization (periodicity) of the signal over the "noisiness" in the voice signal. The CPPS is a modification of the CPP measure, where the individual cepstra are smoothed across time and quefrency domains, which was developed for greater prediction accuracy particularly in speech signals.[2]

The CPP and CPPS measures were shown to be more reliable than the traditional perturbation measures such as jitter, shimmer, and noise to harmonic ratio.[5−7] A higher CPP amplitude value can be found in highly periodic signals and lower CPP amplitude value in less periodic or aperiodic signals.[1,5] From previous clinical studies, CPP and CPPS measures have been found to correlate strongly with perceptual evaluations of voice.[2,6,8] Applications of CPP measures have been extended to the analysis of different phonation and dysphonia types. It has been reported that CPP values are higher for pressed and normal (modal) phonation compared to breathy type of phonations.[9] These findings have been attributed to larger open quotient values of glottal waveform during breathy phonations which lead to increased spectral noise.[9] Wolfe and Martin[10] classified dysphonic patients into breathy, hoarse and strained voice types based on four parameter model including cepstral peak prominence. The CPP values were lower for hoarse and breathy voice compared to strained voice type.[10] Lower CPP values have been reported to differentiate rough from normal voice based on the increased amplitude of noise components in relation to fundamental frequency in rough voice.[11] The CPP measure has also been useful to differentiate hypofunctional from normal voice.[12] Perceptual evaluation of strain severity has as well shown moderate to high correlation with the cepstral measures.[13] CPP and other cepstral based measures have also been reported to be useful in assessing voice quality in various voice disorders,[14] vocal

Accepted for publication September 11, 2018.

From the *Voice Research Laboratory, Department of Biophysics, Faculty of Science, Palacký University Olomouc, Olomouc, Czech Republic; †Speech and Voice Research Laboratory, Faculty of Education, University of Tampere, Tampere, Finland; ‡Ear and Oral Diseases, Department of Phoniatrics, Tampere University Hospital, Tampere, Finland; and the §Department of Otorhinolaryngology and Phoniatrics—Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

Address correspondence and reprint requests to Anne-Maria Laukkanen, Speech and Voice Research Laboratory, Faculty of Education, University of Tampere, Åkerlundinkatu 5, 33100 Tampere, Finland. Jan G Švec, Voice Research Laboratory, Department of Biophysics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic; E-mail: Anne-Maria.Laukkanen@uta.fi Jan.Svec@upol.cz

Journal of Voice, Vol. ■■, No. ■■, pp. 1−11
0892-1997

https://doi.org/10.1016/j.jvoice.2018.09.010

nodules[15] and unilateral vocal fold paralysis.[16] CPPS has been recommended for voice screening purposes as it has a high predictive value for voice disorder status.[17]

CPP as well as CPPS have been used to analyze both sustained vowels and continuous speech samples in assessing dysphonic voices. Hillenbrand and Houde[2] reported that both CPPS and CPP were good predictors of breathiness rating, while CPPS showed slightly better results over CPP for both sustained vowel and continuous speech samples. In a study by Hasanvand et al[18] CPPS and CPP were shown to be significantly reduced in female dysphonic subjects compared to non-dysphonic subjects for both vowel and speech (reading) samples. Comparing dysphonic to non-dysphonic males, the authors showed that CPPS from vowel and speech and CPP from only speech sample were significantly reduced. Authors advocate use of both CPP and CPPS for differentiating dysphonic and non-dysphonic individuals. In another study, Brinca et al[19] reported both CPP and CPPS measures to differentiate between dysphonic and normal individuals for sustained vowel sample, but only CPP from continuous speech sample to help differentiating between the two groups. These authors as well report both CPP and CPPS to be promising acoustic measures of dysphonia. Moers et al[20] reported the reading-based CPP and CPPS to correlate well with perceptual rating of dysphonic voice. Based on all these results, in this paper we explore the use of both CPP and CPPS measures for both vowel and speech samples.

Occupational voice users normally demand more attention than nonoccupational voice users. CPP measures have been applied to assess voice quality in vocally healthy occupational voice users, such as radio broadcasters[21] and in Indian Carnatic classical singers.[22] The Indian Carnatic classical singers had a higher CPP compared to nonsingers, which could be attributed to stronger harmonic organization of voice in the singers.[22] However, in the study on radio broadcasters, there was no difference between the radio performers and nonradio performers on the cepstral measures indicating no differences in the strength of harmonic content in the voice signal between the two groups.[21]

One of the largest groups of professional voice users are teachers. Teacher's voice is vulnerable to disorders as a result of prolonged voice use and heavy vocally loading conditions.[23] Poor environmental[23−26] and working conditions,[27,28] unawareness of appropriate vocal hygiene[29] and lack of voice training,[30] all may contribute to the development of voice disorders in teachers. Several studies have shown a high prevalence of frequently occurring symptoms of vocal overloading and fatigue in teachers.[27,30−32] Studies have shown that in presence of unfavorable environmental conditions such as background noise, teachers tend to raise their voice and speak with increased vocal loudness leading to increased vocal effort and strain in these teachers.[33−35] There have been some indications, that cepstral peak prominence values may be influenced by vocal loudness and depend on the sound pressure level (SPL) of voice.[36] This relationship has not yet been well explored and deserves more attention, however.

The present study applies cepstral (CPP and CPPS) and perceptual evaluations to assess voice quality of female primary school teachers who are serving in a vocally loading profession and have not been seeking for help for any voice problems. These teachers considered themselves to have normal voice, but in some of them pathological findings in the larynx were discovered through laryngoscopy, which did not make it impossible for them to work as a teacher. The questions addressed in this study are: (1) What is the perception of the voice quality and firmness of phonation for the sustained vowel, comfortable and loud speech in teachers who consider themselves to have normal voice? (2) What are the representative CPP, CPPS and SPL values for sustained vowel, comfortable and loud speech in these teachers? (3) How are these CPP and CPPS values related to the measured voice SPLs? (4) In case of laryngeal pathologies, are these perceivable by voice expert listeners and detectable by the CPP, CPPS and voice SPL measures?

MATERIALS AND METHODS

Participants and their laryngeal status

The material for this study was derived from an earlier study,[37] which investigated the relationship between self-reported voice symptoms, working conditions, background factors (such as noise and air quality), and phoniatric evaluation, but did not attempt using cepstral measures in these teachers. A total of 84 Finnish female primary school teachers volunteered as subjects for this study. Their mean age was 42.6 ± 8.9 years. Their mean time in profession was 16.5 ± 9.4 years. The mean number of teaching hours per week was 31.3 ± 7.3 . All the participants considered themselves to be vocally healthy and capable of carrying their profession. Some laryngeal changes were found in 44 (52.4%) teachers; 33 of them (39.3 %) had mild and 11 (13.1 %) had substantial changes that were evaluated by an experienced phoniatrician on a three point scale (1-healthy; 2-mild changes; 3-disordered). This laryngeal status rating was based on case history and indirect mirror laryngoscopy. Mirror laryngoscopy was used out of practical reasons, since the laryngeal inspections were mostly made in field conditions and no portable rigid endoscopy system was available for that purpose. The mild laryngeal changes consisted of mild vocal fold erythema, arytenoid erythema, mild edema, and mild glottal closure insufficiency. The more substantial findings (disordered group) included individuals having nodules, polyps, chronic laryngitis, laryngeal reflux disease, and moderate to severe glottal closure insufficiency.[37] Table 1 lists the laryngeal findings in the participants of the present study diagnosed via indirect laryngoscopy.

Recordings and tasks

Teachers were asked to sustain three times a prolonged vowel [a:] for 5 seconds, followed by reading of a text containing 213 words (no sibilants were included in the text to reduce speech noise components in the signal) at comfortable loudness as in conversational speaking. Additionally, the teachers were asked to read the same text at an increased loudness level as if teaching in a large noisy classroom. The voice recordings were carried out in primary schools, in teacher's own classrooms with minimal ambient noise (approximately about 35 dB(A)). Recordings were made using a portable digital recorder (Sony TCD-D8, Sony Corporation, Tokyo, Japan) and an omnidirectional head-mounted microphone (C477, AKG, Vienna, Austria), selected according to the recommendations by Švec and Granqvist (2010).[38] The microphone was maintained at a constant distance of 6 cm, at an angle of 45° from the side of the subject's mouth. The voice recordings were then calibrated using a sound level meter (type 2206 Brüel & Kjær, Copenhagen, Denmark) to obtain the true SPL of vowel and speech samples.

SPL calibration procedure and measurement

Calibration was made by using a standard complex sound source (BOSS-TU 120), and the sound level meter (SLM), placed at the same distance and angle from the sound source as the microphone was from the subject's lips. For SLM, the slow time averaging and C-frequency weighting was used. After the recording, the sound calibration signal was then loaded in *Praat* software. For calibrating the sound levels in the *Praat*, the procedure mentioned by Boersma and Weenink (2013)[39,40] in the *Praat* manual was used, where the recorded signal was mathematically amplified to obtain the true sound pressure levels (that corresponded to the waveform values in pascals) using the multiplication factor $10^{\left(\frac{\Delta L}{20}\right)}$ where ΔL was the difference level (difference between the true sound pressure level read in the SLM and the uncalibrated level depicted in the *Praat* software). In *Praat* this was done by selecting the signal and choosing the option "Multiply" from the "Modify" menu and supplying the multiplication factor.

After the calibration, the steps involved in obtaining the SPL value in *Praat* were as follows:

Voice sample of interest (vowel or speech) was selected in the "View and edit" window of *Praat*. From the "Intensity settings" the intensity contour was obtained by selecting the option "Show intensity". The following intensity settings were used: view range 40−120 dB, "mean energy" averaging method, and "subtract mean pressure" chosen (as in standard settings). The final representative SPL value was obtained using the "Get intensity" option. The final single SPL value obtained this way represents a close approximation of the time-averaged (equivalent) C-weighted sound level for the entire voice sample selected as measured by the sound level meter.[40] Briefly, the time-averaged sound level of a voice signal is equivalent to SPL of a steady sound which has the same duration and energy as the selected voice signal; C-weighting assures the voice spectrum is minimally influenced within the range of 32−8000 Hz.[40]

TABLE 1.
Diagnostic Distribution of Study Participants. (Some Participants Presented With More Than One Finding)

Laryngeal Status Category (No. of Subjects)	Laryngeal Findings	Number of Subjects
Healthy (40)	Normal laryngeal findings	40
Mild changes (33)	Mild redness of vocal folds (VF)	4
	Mild swelling of VF	5
	Beginning vocal nodule	1
	Mild redness in arytenoids	7
	Slight amount of thick mucus	2
	Slight hoarseness	7
	Incomplete glottal closure in phonation	10
	Mild false VF medialization	2
	Slight hyperkinesia	3
Disordered (11)	Nodules	4
	Polyps	2
	Chronic laryngitis	1
	Vocal fold atrophy	2
	Reflux disease	1
	Moderate to severe closure insufficiency	1

Cepstral analysis

Sustained vowel [a:] at comfortable loudness for 3 seconds and 2 first sentences of continuous speech samples (23 syllables) at comfortable and increased loudness were analyzed for all teachers for CPP and CPPS data using software *Praat*. The vowel samples were chosen from the middle and most stable part of the second vowel from the row of three trials recorded. These selections were identical to those used for SPL analysis. The CPP values were obtained using standard *Praat* (version 5.4.05) settings while the CPPS values were extracted with settings recommended by Maryn and Weenink (2015).[3] Table 2 shows the steps and parameter setting in *Praat* software for the extraction of CPP and CPPS.

Perceptual analysis

The same samples of comfortable vowel phonation and comfortable and loud speech reading that were analyzed for cepstral measures were also perceptually analyzed by five experienced voice raters. They used headphones

TABLE 2.
The Steps and Parameter Setting in the *Praat* Software for Extraction of CPP and CPPS Values for the Vowel and Continuous Speech Samples

Step 1) Select the vowel or speech sample
Step 2) Go to "Analyze periodicity" and click on to "To Power cepstrogram" in the *Praat* Objects window.
Step 3) Use the following settings for generating the power cepstrogram:

Parameter setting	CPP (standard settings for *Praat* version 5.4.05)	CPPS[3]
Pitch floor (Hz)	60	60
Time step (s)	0.002	0.002
Maximum frequency (Hz)	5000	5000
Pre-emphasis from (Hz)	50	50
Step 4) On selecting the newly generated "powercepstrogram" click on to "Query" and select "Get CPPS" from the menu, and use the following settings:		
Select subtract tilt before smoothing	Yes	No
Time averaging window (s)	0.001	0.01
Quefrency averaging window (s)	0.00005	0.001
Peak search pitch range (Hz)	60–330	60–330
Peak search tolerance (0–1)	0.05	0.05
Interpolation	Parabolic	Parabolic
Tilt line quefrency range (s)	0.001–0.0 (=end)	0.001–0.0 (=end)
Line type	Exponential decay	Straight
Fit method	Robust	Robust

(Sony MDR-CD480) in the evaluation task. They rated overall voice quality along a ten point unipolar scale from 0 = poor to excellent = 10. Additionally, they evaluated the vocal firmness along a bipolar axis from 0 = breathy through 5 = adequate to 10 = pressed. The listeners could listen to each sample as many times as they liked in order to be sure of the evaluation. The individual listeners' ratings were averaged for each sample to be used in statistical analyses.

Statistical analyses

Kolmogorov–Smirnov test was used to check normal distribution of voice SPL, cepstral measures (CPP and CPPS) and perceptual ratings (voice quality and firmness) for all the three voice samples. To check the inter-rater reliability for the perceptual ratings, Cronbach's alpha test was used. Paired *t*-test was used to compare voice SPL, CPP, CPPS, voice quality and firmness ratings between (a) comfortable vowel and comfortable speech, and (b) between comfortable speech and loud speech. Pearson's product moment correlation test was used to find correlations between voice SPL and cepstral measures. Spearman's rank order correlation test was used to find correlations of the laryngeal status categories (healthy, mild changes, and disordered) with perceptual ratings and acoustic measures. One way ANOVA was used to compare the voice quality rating, firmness rating, cepstral measures and voice SPLs across the three laryngeal status categories. All the statistical analyses were carried out using SPSS 22 software (IBM SPSS Statistics v. 22 for Windows, Armonk, NY). Significance level was set at $P < 0.05$ in the statistical analyses. MATLAB R2016a was used for scatterplots.

Ethical approval

Permission for data collection was obtained from school administration and social services departments in the districts in question. Participants volunteered in the study and signed a written consent, which informed them about the aim and procedure in the studies, and stated that the participants may withdraw from the study at any point without any consequences. Handling and preservation of the research material follows the Personal Data Act (523/1999) of Finland.

RESULTS

All the measures, CPP, CPPS, SPL, voice quality rating, and rating of firmness for all three voice samples, were normally distributed based on Kolmogorov–Smirnov Test.

Reliability of perceptual evaluation

The inter-rater reliability of the perceptual evaluation (Table 3) was regarded as adequate based on results of Cronbach's alpha except for rating of voice quality for loud speech which was lower (0.60), and was found questionable, as normally the cutoff value of 0.70 is considered acceptable for reliability.[41]

Acoustic and perceptual results for the three voice tasks

The results of the acoustic and perceptual evaluations for the three voice tasks are shown in Table 4. Furthermore,

TABLE 3.
Inter-Rater Reliability for Voice Quality and Firmness Rating for Three Voice Samples

Voice Samples	Inter-Rater Reliability—Cronbach's Alpha (α) Value	
	Vocal Quality	Vocal Firmness
Sustained comfortable vowel	0.83	0.80
Comfortable speech	0.82	0.83
Loud speech	0.60 (low and questionable)	0.82

the results of the paired t-tests evaluating the significance of the differences between the different tasks are shown in Table 5. Together, these tables reveal that: (a) The CPP, CPPS and SPL values were significantly larger for sustained vowels than for speech at comfortable loudness; (b) The CPP, CPPS and SPL values were significantly larger for loud speech than for comfortable speech; (c) Perceptually, the voice quality was found (marginally) significantly better for comfortable vowel than comfortable speech whereas vocal firmness did not show any significant differences here; (d) The voices were found to have significantly better quality and more firmness/less breathiness for loud speech than for comfortable speech.

CPP and CPPS versus SPL for vowel and speech

The next aim was to find the relationship between cepstral and voice SPL measures. Table 6 shows the Pearson's product moment correlation between the cepstral and voice SPL measures. The results show a positive moderate correlation between voice SPL and both CPP and CPPS for vowel. Also a positive moderate correlation was obtained between voice SPL and CPPS for loud speech and a mild correlation with CPP for loud speech. No significant correlations were obtained between voice SPL and cepstral measures for comfortable speech. However, when the comfortable and loud speech data were pooled together, the voice SPL again correlated moderately with both CPP and CPPS measures.

The relationship between the SPL and the cepstral measures is demonstrated more clearly in Figure 1. For sustained vowel the regression line through the data revealed these relationships:

$$CPP = 0.2 * SPL + 3.7 \text{ and } CPPS = 0.18 * SPL - 1.4 \qquad (1a, b)$$

These relationships indicate that for a 10 dB increase in SPL there was, on average, 2.4 dB increase in CPP and 1.8 dB increase in CPPS. Also, when exploring these relationships we may find, e.g., that for the SPL of 80 dB the CPP and CPPS show the average values of 22.9 dB and 13 dB, respectively, for the sustained vowels.

For speech, both comfortable and loud pooled together, linear regression revealed these relationships:

$$CPP = 0.073 * SPL + 13 \text{ and } CPPS = 0.12 * SPL + 1.5 \qquad (2c, d)$$

These relationships indicate that for a 10 dB increase in SPL there was, on average, 0.7 dB increase in CPP and 1.2 dB increase in CPPS. For the SPL of 80 dB the corresponding average CPP and CPPS values of speech were 18.8 dB and 11.1 dB, respectively.

TABLE 4.
The Evaluation Results Expressed Through the Mean and Standard Deviation Values for the three Voice Samples

Voice Samples	CPP (dB)	CPPS (dB)	Voice SPL(dB)	Voice Quality	Vocal Firmness
Sustained vowel	23.4 ± 2.9	13.6 ± 2.1	82.4 ± 5.5	4.7 ± 0.9	5.1 ± 1.0
Comfortable speech	19.0 ± 1.4	10.4 ± 1.5	76.4 ± 3.3	4.4 ± 1.0	4.8 ± 1.2
Loud speech	19.6 ± 1.2	11.4 ± 1.4	84.9 ± 3.8	4.9 ± 1.0	5.7 ± 1.2

TABLE 5.
***P* values for Paired *t*-test Comparing the Evaluation Results for Vowel Versus Speech at Comfortable Loudness and for comfortable Versus Loud Speech. Significant values ($P < 0.05$) are indicated by ***

Voice Samples	CPP	CPPS	SPL	Voice Quality	Vocal Firmness
Vowel versus comfortable speech	$P < 0.001$*	$P < 0.001$*	$P < 0.001$*	$P = 0.040$*	$P = 0.085$ (not significant)
Comfortable speech versus loud speech	$P < 0.001$*	$P < 0.001$*	$P < 0.001$*	$P < 0.001$*	$P < 0.001$*

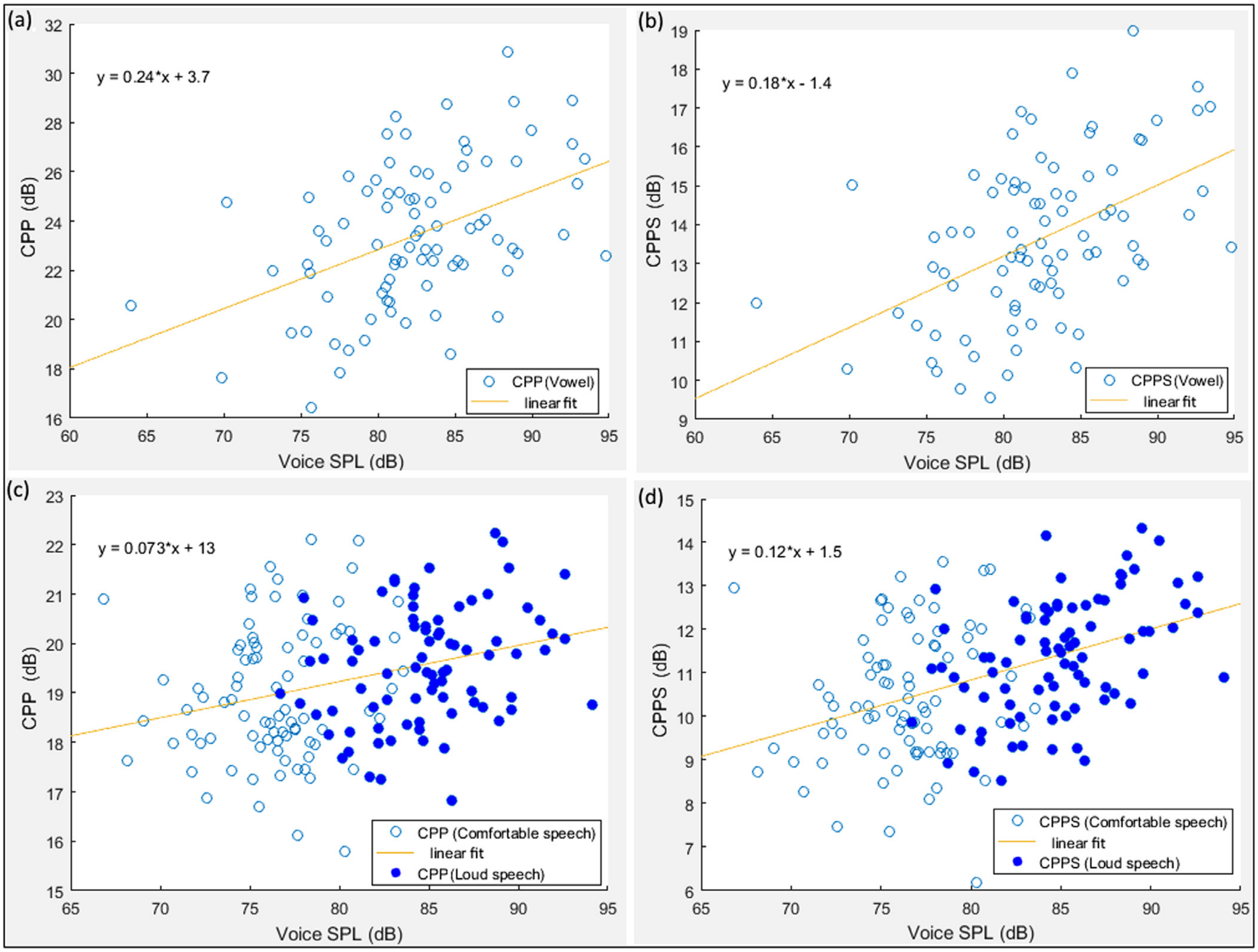


FIGURE 1. A scatterplot showing relationship between time-averaged equivalent SPL (at 6 cm distance in dB re 20 μPa) and the two cepstral measures for sustained vowel [a, b] and speech [c, d]. The speech data contain both the comfortable (empty circles) and loud (filled circles) conditions together. Notice the linear regression lines with their equation shown in each of the graphs—all of them show the trend of CPP/CPPS increase with increased SPL.

CPP/CPPS = cepstral peak prominence/smoothed cepstral peak prominence; SPL = sound pressure level.

Voice perception versus laryngeal pathology

Neither of the perceptual ratings correlated with the laryngeal status findings for any of the vocal tasks according to the Spearman's rank order correlation. Also results of one way ANOVA test showed no significant differences across the three laryngeal status categories ($P > 0.05$) for the perceptual ratings of voice quality and firmness. No systematic trends were found across the different laryngeal status categories either. The numerical results are shown in Table 7.

Acoustic measures (CPP, CPPS, and SPL) versus laryngeal pathology

Similar to the perceptual ratings, none of the acoustic measures (CPP, CPPS, or SPL) correlated with laryngeal status categories, for any of the three voice samples according to the Spearman's rank order correlation test. Also results of ANOVA showed no significant differences for the acoustic measures across the three groups divided on the basis of laryngeal status evaluation ($P > 0.05$). Nevertheless, the data showed a small but systematic decrease of the CPP, CPPS and SPL values from healthy to mild to disordered category in all the three voice samples. The numerical results are shown in Table 8.

DISCUSSION

Sustained vowels and speech at comfortable loudness are standard tasks used in clinical evaluation of voice.[4,42–44] As teachers often use loud speech when teaching, in this study we also added speaking at a raised loudness level as the third voice task. We were interested in finding out how these tasks influence the CPP, CPPS and SPL values and perceptual ratings of vocal quality and firmness in teachers who considered themselves to have normal voice. Furthermore, since in some of the teachers laryngeal pathologies were

TABLE 6.
Pearson's Product Moment Correlation Values and *P* Values for Correlations Between Cepstral Measures and Voice SPL Measures. Significant values (*P* < 0.05) are indicated by *

Voice Samples	CPP versus SPL		CPPS versus SPL	
	Correlation Value (r)	Sig. (2-Tailed)	Correlation Value (r)	Sig. (2-Tailed)
Vowel	0.45	$P < 0.001$*	0.49	$P < 0.001$*
Comfortable speech	0.16	$P = 0.137$	0.18	$P = 0.096$
Loud speech	0.28	$P = 0.009$*	0.45	$P < 0.001$*
Combined comfortable and loud speech	0.31	$P < 0.001$*	0.43	$P < 0.001$*

detected laryngoscopically, it was of interest to find out whether some of these measures could help in detecting these underlying vocal pathologies despite of the fact that they were not self-perceived by the teachers.

CPP and CPPS measures were of particular interest here. These measures started to be explored after Hillenbrand[1,2,45]_ developed the *SpeechTool* software (James Hillenbrand; Western Michigan University, Kalamazoo, MI—https://homepages.wmich.edu/~hillenbr/), for the extraction of the cepstral peak prominence measures from the voice samples. After then, the cepstral measures have been implemented also in other software packages such as *Computerized Speech Lab* (CSL, Kay Pentax, Lincoln Park, NJ) and the freely available *Praat*[3] (Paul Boersma and David Weenink, Institute of Phonetic Sciences—University of Amsterdam, The Netherlands-http://www.praat.org/). Since their implementation in 2015, the cepstral measures in *Praat* have been used in measuring the cepstral values in normophonic and dysphonic individuals.[3,17,46] The authors Maryn and Weenink[3] reported that the CPPS values (for both vowel and continuous speech sample) obtained from *Praat* were a "highly acceptable approximation" of CPPS obtained from *SpeechTool* software.[1,2] Also Sauder et al[17] reported that the smoothed CPP for connected speech samples derived from *Praat* software, had a high rate of accuracy in predicting voice disorder status with excellent sensitivity value of 90% on the area under the receiver operating characteristic curve. Nevertheless, CPP and CPPS results obtained from the same voice samples using different software packages yield different absolute values.[46−48] In this study we measured the CPP and CPPS values extracted from *Praat* software and related them to SPL and perceptual evaluations of voice quality and vocal firmness.

From Table 4 and 5 we can observe that the CPP and CPPS values were significantly larger (on average by 4.4 dB and 3.2 dB, respectively) for comfortable sustained vowels compared to comfortable speech. This is an expected result, because speech contains fundamental frequency and intensity fluctuations, voice onsets and offsets, vocal pauses, etc., all of which decrease the prominence of harmonic organization over the noise content measured by the CPP and CPPS parameters.[49,50] The SPL of the vowel at comfortable loudness was also, on average, 6 dB larger than that of speech at comfortable loudness. This can be attributed to voiceless consonants, and pauses between words and sentences, all of which decrease the average sound level of the speech sample compared to sustained vowels.

As far as the loud versus comfortable speech comparisons are concerned, Table 4 reveals that the SPL increased on average by 8.5 dB from comfortable to loud speech and simultaneously the CPP and CPPS values increased on average by 0.6 and 1 dB, respectively. This significant trend of increasing CPP and CPPS values with increased SPL is more explicitly shown through the regression lines in Figure 1 which are quantified and mathematically expressed through Equations (1) and (2). The cepstral prominence dependencies on SPL are slightly different for vowels than for speech, but the trend is the same in both vocal tasks. This relationship confirms the previous findings that the cepstral measures increase with increasing SPL of voice[36] and is consistent with other studies reporting improvements of perturbation measures with increased voice intensity. For speech at comfortable loudness, there was no significant correlation between SPL and CPP or CPPS (Table 6). This can be attributed to a smaller range of SPL observed in this vocal task in combination with the rather large spread of CPP and CPPS values for the different individuals (recall Figure 1c and d, data indicated by empty circles only). However, the correlations between SPL and CPP or CPPS became highly significant ($P < 0.001$) when the SPL range was enlarged by pooling the comfortable and loud speech conditions together (both empty and filled circles in Figure 1c and d). The relationship between voice SPL and cepstral peak prominence values may be related to previous findings that increased phonational loudness decreases the perturbation in voice[51−53]thus leading to increase in CPP/CPPS values of vowel phonation.[36] This has been related to an increase in the medial compression of vocal folds that improves the glottal closure, decreases glottal noise, and increases the strength of overtones in the signal.[36,54] The assumption of increased medial compression of the vocal folds is supported here by the perceptually increased firmness in loud voice, as shown in Figure 1.

It is interesting to compare our CPP and CPPS results to those found in other studies for healthy and disordered subjects. Here, the CPP and CPPS settings also need to be considered. Our CPPS measurement procedure and settings were identical to those specified for measuring the Acoustic

TABLE 7.
Mean, Standard Deviation (SD) and Standard Error (SE) of Perceptual Ratings for all Three Voice Samples for Teachers Grouped Under the Laryngeal Status Category

Laryngeal Status Category	N	Voice Quality Rating			Firmness Rating		
		Vowel	Comfortable Speech	Loud Speech	Vowel	Comfortable Speech	Loud Speech
		Mean ± SD (SE)	Mean ± SD (SE)	Mean ± SD (SE)	Mean ± SD (SE)	Mean ± SD (SE)	Mean ± SD (SE)
Healthy	40	4.7 ± 0.9 (0.14)	4.1 ± 1.0 (0.15)	4.9 ± 0.9 (0.14)	5.1 ± 0.9 (0.14)	4.6 ± 1.3 (0.20)	5.7 ± 1.3 (0.21)
Mild	33	4.8 ± 0.9 (0.15)	4.7 ± 1.1 (0.18)	5.0 ± 1.0 (0.17)	5.2 ± 1.0 (0.17)	4.9 ± 1.1 (0.18)	5.8 ± 1.1 (0.19)
Disordered	11	4.4 ± 0.6 (0.18)	4.5 ± 0.9 (0.26)	4.9 ± 1.0 (0.30)	4.6 ± 0.6 (0.18)	4.9 ± 1.4 (0.43)	5.6 ± 1.1 (0.32)

Voice Quality Index (AVQI) using *Praat* software.[3] The mean CPPS values for all the teachers in this study (vowel: 13.6 ± 2.1 dB; comfortable speech: 10.4 ± 1.5 dB; and loud speech: 11.4 ± 1.4 dB) are similar to the results of CPPS obtained by Maryn and Weenink (2015)[3] using *Praat* software on a group of 289 normal and dysphonic individuals). They report a CPPS value of 11.66 ± 2.68 dB for a concatenated voice sample which combines 3- seconds-long vowel sample and connected speech together in one file.[3,55] Also, our CPPS results are in good correspondence with those obtained by Latoszek et al (2017)[56] using the AVQI-based CPPS setup (*Praat* version 5.3.57), where they report the mean CPPS value of 11.92 ± 2.15 dB in individuals with perceptually nondysphonic voices.

On the other hand, our results for CPPS are more than 9 dB different from those obtained by Sauder et al (2017),[17] where they report a value of 20.11 ± 1.27 dB for non-dysphonic subjects on continuous speech sample using the default *Praat* (version 6.0.17) settings. These values, however, are close to our CPP speech values (19.0 ± 1.4 dB for comfortable and 19.6 ± 1.2 dB for loud speech). Similarly, the CPPS values reported by Watts et al (2017)[46] on vowel and speech (i.e., 22.86 ± 4.07 dB for English vowel and 20.07 ± 3.33 dB for English sentence) on a group of 22 dysphonic and 22 non-dysphonic speakers are much closer to our CPP values than to our CPPS values. We therefore suspect that the CPPS values reported by Sauder et al (2017)[17] and Watts et al (2017)[46] may not be the smoothed CPPS values but rather the nonsmoothed CPP ones as we have discovered that the current default *Praat* CPPS settings use the time averaging window of 0.001 seconds and quefrency averaging window of 0.00005 seconds that are so short that effectively no smoothing takes place. Different *Praat* versions provide different default settings, and therefore one needs to be cautious in selecting the proper parameter settings for smoothing. Hence, we have listed the exact settings used for this study in Table 2 to assure better comparability and reproducibility of the results.

Different settings and methodology may also explain differences of CPPS values among different software packages. Similarly to the findings of Maryn and Weenink[3] our *Praat*-based CPPS results do not match the values obtained from Hillenbrand's *SpeechTool* software in various other studies[19,20,22,57,58] due to differences in the algorithms used in these softwares. For example Balasubramanium et al[58] reported the mean CPP value of 13.65 ± 0.9 dB and mean CPPS value of 6.30 ± 0.35 dB for vowel sample on 22 normal subjects using *SpeechTool* software and Heman-Ackah et al[57] reported a mean CPPS value of 4.77 ± 0.97 dB in 30 normal voices on a running speech sample. These values are much lower than ours. Therefore, choosing appropriate software and accurate parameter settings is an important consideration when performing cepstral analysis of voice. There is a need to unify and standardize the CPP and CPPS measurement procedures in future so that the data are reproducible and better comparable.

Since CPP and CPPS measures have been reported to correlate with perceptual ratings of voice quality,[2,6,8]perceptual evaluations were also included in this study. The inter-rater reliability of the vocal quality and firmness (Table 3) was adequate for all the voice samples except for the voice quality rating of loud speech sample which had low and questionable alpha value. This suggests that perceptual evaluation of quality for loud speech is likely more complex than for vowels or speech at comfortable loudness. Tables 4 and 5 reveal that the voice quality was slightly better for vowel than speech in comfortable loudness. This may be due to the fact that speech samples are more demanding on laryngeal coordination and expose voice abnormalities more extensively than sustained vowels.[59] However, the voice quality values approached value of 5 (on the rating scale of 0 = poor to excellent = 10) for both vowel and comfortable speech samples indicating, on average, good voice quality in these teachers. The vocal firmness did not show any significant differences between sustained vowels and comfortable speech and their values as well approached normal values along the continuum from breathy (0)

TABLE 8.
Mean, Standard Deviation (SD) and Standard Error (SE) of Cepstral Values (in dB) and of the Time-Averaged Equivalent SPL at 6 cm Distance (in dB re 20 μPa) for All Three Voice Samples for Teachers Grouped Under the Laryngeal Status Category

Laryngeal Status Category	N	Vowel			Comfortable Speech			Loud Speech		
		CPP Mean ±SD (SE)	CPPS Mean ±SD (SE)	SPL Mean ±SD (SE)	CPP Mean ±SD (SE)	CPPS Mean ±SD (SE)	SPL Mean ±SD (SE)	CPP Mean ±SD (SE)	CPPS Mean ±SD (SE)	SPL Mean ±SD (SE)
Healthy	40	23.8 ±2.6 (0.41)	13.9± 1.9 (0.30)	83.2 ±6.1 (0.97)	19.0 ±1.5 (0.23)	10.5 ±1.2 (0.25)	77.0 ±3.0 (0.47)	19.7 ±1.3 (0.20)	11.5 ±1.5 (0.23)	85.0 ±3.8 (0.60)
Mild	33	23.0 ±3.4 (0.59)	13.4 ±2.4 (0.41)	81.8 ±5.2 (0.90)	19.0 ±1.2 (0.21)	10.4 ±1.5 (0.26)	76.2 ±3.9 (0.68)	19.5 ±1.1 (0.18)	11.3 ±1.2 (0.21)	85.0 ±4.1 (0.70)
Disordered	11	23.0 ±2.3 (0.68)	13.3 ±1.6 (0.49)	81.0± 3.5 (1.06)	18.8 ±1.4 (0.43)	10.1 ±1.5 (0.44)	74.9 ±2.1 (0.62)	19.4 ±1.4 (0.32)	11.2 ±1.2 (0.38)	84.0 ±2.9 (0.88)

through normal (5) to pressed (10), suggesting normal vocal fold adduction in these teachers.

Despite of the reduced reliability of voice quality evaluations for loud voice found here, the perceptual differences were much more prominent between the comfortable and loud speech samples than between comfortable vowels and comfortable speech. The voices were found to have significantly better quality and more firmness/less breathiness for loud speech. This can be again related to the reduction of voice perturbations in louder speech.[36,51] The mean values for firmness ratings in loud speech samples (5.7 ± 1.2, Table 4) suggest that the voices were neither breathy nor pressed, suggesting that the teachers on average did not have the tendency to endanger their larynges by inadequate voice production mechanisms. This may be related to the fact that the teachers had no vocal complaints.

Despite of no vocal complaints, laryngeal pathologies were detected laryngoscopically in some of the teachers. Another goal of the present study was therefore to find out whether the expert perceptual evaluations, cepstral measures, and voice SPL measures could reveal the potential vocal changes due to the underlying pathology. The results of one way ANOVA test did not show any systematic differences for perceptual ratings across the laryngeal status categories (healthy, mild changes, disordered), and no significant correlations were found between the perceptual and laryngeal status evaluation. From Table 7 it can be seen that the voice quality and firmness ratings do not show any specific trends. This indicates that the laryngeal pathologies were not well perceivable by the voice expert listeners.

Similar to perceptual evaluations, also the acoustic measures did not show statistically significant differences among or correlations with the laryngeal status categories for any of the vocal tasks. Nevertheless, a closer look at the results in Table 8 revealed that, in contrast to the perceptual evaluations, the cepstral measures (both CPP and CPPS) and SPL values for all three voice samples, show a consistent decline in the mean values with increased severity of the laryngeal pathology. This suggests that the cepstral and SPL measures could be more sensitive to the underlying vocal pathology than the perceptual measures. However, the differences among the disordered versus nondisordered groups were only around 0.2 dB for CPP and CPPS and 2 dB for SPL (Table 8). These differences were much smaller compared to the standard deviations which were above 1 dB for CPP and CPPS and above 2.9 dB for SPL within each category and thus not significant. This indicates that the CPP, CPPS and SPL variability among healthy larynges was larger than the potential influence of the underlying laryngeal pathology in our teachers.

Considering this and the fact that all the teachers considered themselves to have a normal voice, they may be as such referred to having a *functionally healthy voice.* The large variability with respect to the small effect of the laryngeal pathology limits the possibility of using solely the CPP, CPPS and SPL measures for detecting the laryngeal pathology in individual teachers without vocal complaints.

Nevertheless, the trend of CPP, CPPS and SPL lowering with underlying laryngeal pathology can be explored in future for detecting differences among the groups of pathologic and control subjects. In this study, the pathologic group size was limited to only 11 teachers causing the standard error of the mean to be rather large for finding significant differences. Future studies may explore the differences with a larger number of subjects in which the standard error of the mean is expected to be smaller thus revealing better on potential significant differences among different subject groups.

CONCLUSION

The present study brings basic information on CPP and CPPS values in teachers without vocal complaints and their relationships to voice SPL, voice quality, firmness of voice, and underlying laryngeal pathologies. The results show that with increased loudness and SPL the cepstral values increased and the voice became firmer without becoming excessively pressed. Although the teachers considered themselves vocally healthy, 52.4% of them had some laryngeal changes detected laryngoscopically. These underlying pathologies, however, did not significantly correlate with any of the acoustic measures nor with the perceptual judgments of voice quality and firmness confirming the self-perception of the teachers that their voices were functionally healthy. Nevertheless, the cepstral measures and voice SPLs showed a consistent decline in their values with increased severity of laryngeal pathology. This trend may further be explored in future studies.

Acknowledgments

The study was accomplished during the research stay of Ketaki Vasant Phadke at the University of Helsinki. Her stay was supported by the funds from Erasmus Plus studies within program countries (No. 2017−2018/164). The research of the Czech authors (KVP and JGS) has been supported by the Czech Science Foundation (Grantová Agentura České Republiky—GAČR) project GA16-01246S.

REFERENCES

1. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech, Lang Hearing Res*. 1994;37:769–778. http://dx.doi.org/10.1044/jshr.3704.769.
2. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech, Lang Hearing Res*. 1996;39:311–321. http://dx.doi.org/10.1044/jshr.3902.311.
3. Maryn Y, Weenink D. Objective dysphonia measures in the program praat: smoothed cepstral peak prominence and acoustic voice quality index; *J Voice*. 2015;29:35–43. https://doi.org/10.1016/j.jvoice.2014.06.015.
4. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech-Lang Pathol*. 2018;27:887-905. http://dx.doi.org/10.1044/2018_AJSLP-17-0009.
5. Heman-Ackah YD, Michael DD, Baroody MM, et al. Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003;112:324−333. https://doi.org/10.1177/000348940311200406.
6. Heman-Ackah YD, Michael DD, Goding GS. The relationship between cepstral peak prominence and selected parameters of dysphonia; *J Voice*. 2002;16:20–27. https://doi.org/10.1016/S0892-1997(02)00067-X.
7. Leong K, Hawkshaw MJ, Dentchev D, et al. Reliability of objective voice measures of normal speaking voices; *J Voice*. 2013;27:170–176. https://doi.org/10.1016/j.jvoice.2012.07.005.
8. Awan SN, Roy N, Jette ME, et al. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V. *Clin Linguist Phonetics*. 2010;24:742–758. http://dx.doi.org/10.3109/02699206.2010.492446.
9. Shue YL, Chen G, Alwan A. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. In: Eleventh Annual Conference of the International Speech Communication Association INTERSPEECH. 2010: pp. 34−37. https://www.isca-speech.org/archive/interspeech_2010/i10_0034.html.
10. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity; *J Commun Disord*. 1997;30:403–416. https://doi.org/10.1016/S0021-9924(96)00112-8.
11. Awan SN, Roy N. Acoustic prediction of voice type in women with functional dysphonia; *J Voice*. 2005;19:268–282. https://doi.org/10.1016/j.jvoice.2004.03.005.
12. Watts CR, Awan SN. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *J Speech, Lang Hearing Res*. 2011;54:1525–1537. http://dx.doi.org/10.1044/1092-4388(2011/10-0209).
13. Lowell SY, Kelley RT, Awan SN, et al. Spectral-and cepstral-based acoustic features of dysphonic, strained voice quality; *Ann Otol Rhinol Laryngol*. 2012;121:539–548. https://doi.org/10.1177/000348941212100808.
14. Zieger K, Schneider C, Gerull G, et al. Cepstrum analysis in voice disorders. *Folia Phoniatr*. 1995;47:210–217. http://dx.doi.org/10.1159/000266352.
15. Radish Kumar B, Bhat JS, Prasad N. Cepstral analysis of voice in persons with vocal nodules; *J Voice*. 2010;24:651–653. https://doi.org/10.1016/j.jvoice.2009.07.008.
16. Balasubramanium RK, Bhat JS, Fahim S, et al. Cepstral analysis of voice in unilateral adductor vocal fold palsy; *J Voice*. 2011;25:326–329. https://doi.org/10.1016/j.jvoice.2009.12.010.
17. Sauder C, Bretl M, Eadie T. Predicting voice disorder status from smoothed measures of cepstral peak prominence using praat and analysis of dysphonia in speech and voice (ADSV); *J Voice*. 2017;31:557–566. https://doi.org/10.1016/j.jvoice.2017.01.006.
18. Hasanvand A, Salehi A, Ebrahimipour M. A cepstral analysis of normal and pathologic voice qualities in Iranian adults: a comparative study; *J Voice*. 2017;31:508.e17–508.e23. https://doi.org/10.1016/j.jvoice.2016.10.017.
19. Brinca LF, Batista AP, Tavares AI, et al. Use of cepstral analyses for differentiating normal from dysphonic voices: a comparative study of connected speech versus sustained vowel in European Portuguese female speakers; *J Voice*. 2014;28:282–286. https://doi.org/10.1016/j.jvoice.2013.10.001.
20. Moers C, Möbius B, Rosanowski F, et al. Vowel-and text-based cepstral analysis of chronic hoarseness; *J Voice*. 2012;26:416–424. https://doi.org/10.1016/j.jvoice.2011.05.001.
21. Warhurst S, McCabe P, Yiu E, et al. Acoustic characteristics of male commercial and public radio broadcast voices; *J Voice*. 2013;27:655.e1–655.e7. https://doi.org/10.1016/j.jvoice.2013.04.012.
22. Balasubramanium RK, Shastry A, Singh M, et al. Cepstral characteristics of voice in Indian female classical Carnatic singers; *J Voice*. 2015;29:693–695. https://doi.org/10.1016/j.jvoice.2015.01.002.
23. Vilkman E. Voice problems at work: a challenge for occupational safety and health arrangement; *Folia Phoniatr*. 2000;52:120–125. https://doi.org/10.1159/000021519.

24. Rantala LM, Hakala S, Holmqvist S, et al. Classroom noise and teachers' voice production. *J Speech, Lang Hearing Res*. 2015;58:1397–1406. http://dx.doi.org/10.1044/2015_JSLHR-S-14-0248.
25. Cutiva LCC, Puglisi GE, Astolfi A, et al. Four-day follow-up study on the self-reported voice condition and noise condition of teachers: relationship between vocal parameters and classroom acoustics; *J Voice*. 2017;31:120.e1–120.e8. https://doi.org/10.1016/j.jvoice.2016.02.017.
26. Durup N, Shield BM, Dance S, et al. Teachers' voice parameters and classroom acoustics—a field study and online survey; *J Acoust Soc Am*. 2017;141. 3540−3540; https://doi.org/10.1121/1.4987482.
27. Kankare E, Geneid A, Laukkanen A-M, et al. Subjective evaluation of voice and working conditions and phoniatric examination in kindergarten teachers; *Folia Phoniatr*. 2012;64:12–19. https://doi.org/10.1159/000328643.
28. Cutiva LCC, Vogel I, Burdorf A. Voice disorders in teachers and their associations with work-related factors: a systematic review; *J Commun Disord*. 2013;46:143–155. https://doi.org/10.1016/j.jcomdis.2013.01.001.
29. Bolbol SA, Zalat MM, Hammam RA, et al. Risk factors of voice disorders and impact of vocal hygiene awareness program among teachers in public schools in Egypt; *J Voice*. 2017;31:251.e9–251.e16. https://doi.org/10.1016/j.jvoice.2016.07.010.
30. Ilomäki I, Mäki E, Laukkanen AM. Vocal symptoms among teachers with and without voice education; *Logopedics Phoniatr Vocol*. 2005;30:171–174. https://doi.org/10.1080/14015430500294106.
31. Simberg S, Sala E, Vehmas K, et al. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period; *J Voice*. 2005;19:95–102. https://doi.org/10.1016/j.jvoice.2004.02.009.
32. Sala E, Laine A, Simberg S, et al. The prevalence of voice disorders among day care center teachers compared with nurses: a questionnaire and clinical study; *J Voice*. 2001;15:413–423. https://doi.org/10.1016/S0892-1997(01)00042-X.
33. Södersten M, Granqvist S, Hammarberg B, et al. Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural DAT recordings; *J Voice*. 2002;16:356–371. https://doi.org/10.1016/S0892-1997(02)00107-8.
34. Phadke KV, Abo-Hasseba A, Švec JG, et al. Influence of noise resulting from the location and conditions of classrooms and schools in Upper Egypt on teachers' voices; *J Voice* 2018;https://doi.org/10.1016/j.jvoice.2018.03.003.
35. Abo-Hasseba A, Waaramaa T, Alku P, et al. Difference in voice problems and noise reports between teachers of public and private schools in Upper Egypt; *J Voice*. 2017;31. 508. e511−508. e516; https://doi.org/10.1016/j.jvoice.2016.10.016.
36. Awan SN, Giovinco A, Owens J. Effects of vocal intensity and vowel type on cepstral analysis of voice; *J Voice*. 2012;26. 670.e15−670.e20; https://doi.org/10.1016/j.jvoice.2011.12.001.
37. Ilomäki I, Leppänen K, Kleemola L, et al. Relationships between self-evaluations of voice and working conditions, background factors, and phoniatric findings in female teachers; *Logopedics Phoniatr Vocol*. 2009;34:20–31. https://doi.org/10.1080/14015430802042013.
38. Švec JG, Granqvist S. Guidelines for selecting microphones for human voice production research. *Am J Speech-Lang Pathol*. 2010;19:356–368. http://dx.doi.org/10.1044/1058-0360(2010/09-0091).
39. Boersma P, Weenink D. *Praat: Doing Phonetics by Computer*; Amsterdam, the Netherlands: Institute of Phonetic Sciences, University of Amsterdam; 2013. http://www.fon.hum.uva.nl/praat/manual/sound_-pressure_calibration.html.
40. Švec JG, Granqvist S. Tutorial and guidelines on measurement of sound pressure level in voice and speech. *J Speech, Lang Hearing Res*. 2018;61:441–461. http://dx.doi.org/10.1044/2017_JSLHR-S-17-0095.
41. Santos JRA. Cronbach's alpha: a tool for assessing the reliability of scales. *J Extension*. 1999;37:1–5.
42. Oates J. Auditory-perceptual evaluation of disordered voice quality; *Folia Phoniatr*. 2009;61:49–56. https://doi.org/10.1159/000200768.
43. Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art; *Auris, Nasus, Larynx*. 2015;42:183–188. https://doi.org/10.1016/j.anl.2014.11.001.
44. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques; *Eur Arch Oto-Rhino-Laryngol*. 2001;258:77–82. https://doi.org/10.1007/s004050000299.
45. Hillenbrand, J.S. Version 1.56 [computer program], 2006.
46. Watts CR, Awan SN, Maryn Y. A comparison of cepstral peak prominence measures from two acoustic analysis programs; *J Voice*. 2017;31:387.e1–387.e10. https://doi.org/10.1016/j.jvoice.2016.09.012.
47. Kim G, Lee Y, Park H, et al. A study of cepstral peak prominence characteristics in ADSV, SpeechTool and Praat. *J Speech-Lang Hearing Disord*. 2017;26:99–111.
48. Madill C, Nguyen DD, Eastwood C, et al. Comparison of cepstral peak prominence measures using the ADSV, speechtool, and voice sauce acoustic analysis programs in vocally healthy female speakers; *Acoust Aust*. 2018;46:215–226. https://doi.org/10.1007/s40857-018-0139-6.
49. Zhang Y, Jiang JJ. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies; *J Voice*. 2008;22:1–9. https://doi.org/10.1016/j.jvoice.2006.08.003.
50. Maryn Y, Roy N. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity; *Jornal da Sociedade Brasileira de Fonoaudiologia*. 2012;24:107–112. http://dx.doi.org/10.1590/S2179-64912012000200003.
51. Brockmann M, Storck C, Carding PN, et al. Voice loudness and gender effects on jitter and shimmer in healthy adults. *J Speech, Lang Hearing Res*. 2008;51:1152–1160. http://dx.doi.org/10.1044/1092-4388(2008/06-0208).
52. Brockmann-Bauser M, Bohlender J, Mehta D. Acoustic perturbation measures improve with increasing vocal intensity in individuals with and without voice disorders; *J Voice*. 2018;32:162–168. https://doi.org/10.1016/j.jvoice.2017.04.008.
53. Brockmann M, Drinnan MJ, Storck C, et al. Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task; *J Voice*. 2011;25:44–53. https://doi.org/10.1016/j.jvoice.2009.07.002.
54. Sulter A, Albers F. The effects of frequency and intensity level on glottal closure in normal subjects; *Clin Otolaryngol Allied Sci*. 1996;21:324–327. https://doi.org/10.1111/j.1365-2273.1996.tb01079.x.
55. Maryn Y, Corthals P, Van Cauwenberge P, et al. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels; *J Voice*. 2010;24:540–555. https://doi.org/10.1016/j.jvoice.2008.12.014.
56. van Latoszek BB, De Bodt M, Gerrits E, et al. The exploration of an objective model for roughness with several acoustic markers; *J Voice*. 2018;32:149–161. https://doi.org/10.1016/j.jvoice.2017.04.017.
57. Heman-Ackah YD, Sataloff RT, Laureyns G, et al. Quantifying the cepstral peak prominence, a measure of dysphonia; *J Voice*. 2014;28:783–788. https://doi.org/10.1016/j.jvoice.2014.05.005.
58. Balasubramanium RK, Karuppali S, Bajaj G, et al. Acoustic-perceptual correlates of voice in Indian Hindu Purohits; *J Voice* 2018;https://doi.org/10.1016/j.jvoice.2018.03.006.
59. Law T, Kim JH, Lee KY, et al. Comparison of rater's reliability on perceptual evaluation of different types of voice sample; *J Voice*. 2012;26. 666. e613−666. e621; https://doi.org/10.1016/j.jvoice.2011.08.003.

Supplement D: Manuscript IV

Influence of Noise Resulting From the Location and Conditions of Classrooms and Schools in Upper Egypt on Teachers' Voices

Authors and their contribution to the study in percentages:

K.V. Phadke 50 %
A. Abo-Hasseba 10 %
J.G. Švec 15 %
A. Geneid 25 %

Journal name and Impact factor (IF):

Journal of Voice, IF (2016): 1.381

Authors' contribution:

K.V. Phadke: Designed the study, analyzed the questionnaire results, applied statistics, and wrote the manuscript.

A. Abo-Hasseba: Designed the questionnaire, collected the questionnaire data on voice of teachers in schools, and approved the final version of the manuscript.

J.G. Švec: Critically revised the successive versions of the manuscripts.

A. Geneid: Organized and supervised the stay of K.V. Phadke in Finland and her participation in the study, designed the questionnaire, helped in designing the study, and revised the manuscript.

Influence of Noise Resulting From the Location and Conditions of Classrooms and Schools in Upper Egypt on Teachers' Voices

*Ketaki Vasant Phadke, †Ahmed Abo-Hasseba, *Jan G. Švec, and ‡Ahmed Geneid, *Olomouc, Czech Republic, †Minia, Egypt, and ‡Helsinki, Finland

Summary: Purpose. Teachers are professional voice users, always at high risk of developing voice disorders due to high vocal demand and unfavorable environmental conditions. This study aimed at identifying possible correlations between teachers' voice symptoms and their perception of noise, the location of schools, as well as the location and conditions of their classrooms.

Method. One hundred forty teachers (ages 21–56) from schools in Upper Egypt participated in this study. They filled out a questionnaire including questions about the severity and frequency of their voice symptoms, noise perception, and the location and conditions of their schools and classrooms. Questionnaire responses were statistically analyzed to identify possible correlations.

Results. There were significant correlations ($P < 0.05$) between voice symptoms, teachers' noise perception, and noise resulting from the location and conditions of schools and classrooms. Teachers experienced severe dysphonia, neck pain, and increased vocal effort with weekly or daily recurrence. Among the teachers who participated in the study, 24.2% felt they were always in a noisy environment, with 51.4% of the total participants reporting having to raise their voices. The most common sources of noise were from student activities and talking in the teachers' own classrooms (61.4%), noise from adjacent classrooms (52.9%), and road traffic (40.7%).

Conclusions. Adverse effect on teachers' voices due to noise from poor school and classroom conditions necessitates solutions for the future improvement of conditions in Egyptian schools. This study may help future studies that focus on developing guidelines for the better planning of Egyptian schools in terms of improved infrastructure and architecture, thus considering the general and vocal health of teachers.

Key Words: Teachers–Voice symptoms–Noise–School and classroom location–Classroom conditions.

INTRODUCTION

Noise, an unwanted sound occurring in the environment in which people live and work,[1] may predispose individuals to have a sense of annoyance or a negative evaluation regarding their environmental conditions.[2] In the past few decades, there have been several research studies conducted to better understand the effects of internal and external noise that prevails in schools and classrooms on student and teacher health, particularly on teachers' vocal conditions.[3–8] There are many sources of background noise inside and outside of classrooms. The predominant outdoor noises include those from automobiles, aircraft, road traffic,[9,10] industrial plants, and activities from school yards and grounds.[11] Indoor noises (also inside classrooms) are primarily from student activities and talking and noise from hallways during breaks between lessons.[9,12] Noise generated within school buildings, including those due to utilities (such as ventilation systems for heating/cooling) often intrudes inside classrooms from walls and partitions, and floor-to-ceiling assemblies.[11]

Accepted for publication March 8, 2018.

The research of the Czech authors (KVP and JGS) has been supported by the Czech Science Foundation (Grantová Agentura České Republiky - GACR) project GA16-01246S.

From the *Department of Biophysics, Faculty of Science, Palacký University Olomouc, Olomouc, Czech Republic; †Department of Otorhinolaryngology—Phoniatric Unit, Faculty of Medicine, Minia University, Minia, Egypt; and the ‡Department of Ear, Nose and Throat—Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

Address correspondence and reprint requests to Ketaki Vasant Phadke, Palacký University Olomouc, Faculty of Science, Department of Biophysics, Voice Research Lab, 17. Listopadu 12, 771 46 Olomouc, Czech Republic. E-mail: ketakislp@gmail.com

Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■

0892-1997

https://doi.org/10.1016/j.jvoice.2018.03.003

A good voice quality is essential for good communication with students, which can otherwise be hampered if teachers experience a voice problem. It has been reported that the vocal impairment of teachers (irrespective of whether it is a mild or severe voice problem) may have some detrimental effects on children's speech-processing ability, resulting in a negative educational effect.[13] An active classroom involves students and teachers conversing at least 60% of the time, pressing the need for a favorable listening environment that supports clear communication.[14] It was found that 13% of the active school teachers in southern Sweden self-reported voice problems.[15] A further serious consequence of this could be job dissatisfaction, disinterest in continuing the job, lack of self-esteem, and fatigue after work.[16] Vilkman has indicated "bad classroom acoustics" to be one of the threats to voice health.[16] In a study by Cutiva et al,[17] the authors systematically reviewed 23 publications and found that most of the studies reported teachers being at high risk of developing voice problems due to noisy classrooms. Noisy classrooms may cause teachers to raise their voices, leading to increased teacher stress and vocal fatigue.[18] Classroom acoustics are often overlooked, where noise, echoes, and reverberation typically interfere with the ability of the listeners to understand speech, thus increasing the vocal effort by teachers.[19] The influence of classroom acoustics on the vocal load of teachers has been documented objectively in recent studies.[6,20–27]

The prevalence of voice disorders among Egyptian teachers is not in the limelight and is an overlooked matter. However, in a recent comprehensive study,[28] authors tried to investigate risk factors for voice disorders in Egyptian teachers of public schools.

The dominant risk factors for poor vocal health in these teachers were overcrowded, loud, noisy, and misbehaved classroom environments, with the teachers having a poor awareness of vocal hygiene. In another Egyptian study,[29] authors compared self-reported voice symptoms and noise reports between public and private schools. They found that teachers working in public schools had more negative voice impacts and were susceptible to their voices failing by the end of their work day, which was also attributed to a larger number of students in the classroom and increased noise disturbances from nearby classes.

Over the years, Egyptian schools have been facing serious issues related to the planning of educational, technical, and architectural requirements.[30] Egyptian school buildings are constructed with inexpensive, local materials, leading to poor building conditions, especially in public schools.[31,32] Moreover, the distribution of school buildings does not follow any standard rule. Some schools are located close to other schools, while some districts in Egypt have no schools to serve the population at all. Based on a survey report, it is mentioned that most of rural Egypt has very few schools and lacks good infrastructure.[33] The location of schools is another problem in Egyptian cities. Most of the schools are located on main streets, near road traffic, or near railway lines, which is a vital factor that requires consideration.[34–37] All of these factors tend to increase noise levels in schools and classrooms, making it an extremely unfavorable working place that affects both teachers' vocal health and children's learning abilities.

The present study is a continuation of the study conducted by Abo-Hasseba et al.[29] Here we hypothesize that the location of a school and its classroom as well as classroom conditions may be a source of noise capable of harming teachers' voices, resulting in a hindrance to their teaching. We sought to identify if any correlations exist between the prevailing voice symptoms of teachers and the noise due to the inappropriate location and conditions of schools and classrooms.

AIM

The study aimed at identifying possible correlations between teachers' voice symptoms and their perception of noise, the location of schools, as well as the location and conditions of their classrooms.

MATERIALS AND METHOD

Participants

Of 200 invited teachers, 140 (85 females and 55 males) between ages 21 and 56 years (mean age = 35.8 years) from schools in Upper Egypt (Governorate of El-Minia) participated in this study. The schools were randomly selected, including both primary and preparatory grades, as well as both public and private school types. Of the 69 teachers from primary schools, 36 taught in public schools and 33 in private. Seventy-one teachers worked in preparatory schools, 34 of which taught in public schools and 37 in private schools. Teachers working in the public and private schools had an average of 17.9 and 7.4 years of teaching experience, respectively, with a total average of 12.3 years combining both groups.[29] Written consent was obtained from all teachers participating in the study.

Questionnaire

Teachers were asked to fill out a questionnaire regarding their demographic data, and the frequency and severity of voice symptoms (dysphonia, laryngeal pain, throat clearing, throat dryness, voice interrupted by the end of the day, and extra voice effort required to continue speaking) from the past 6 months. They were also asked about their school and classroom location and classroom conditions. The teachers also reported their perception on existing noise (and also the source of noise) at their workplace and how they felt about it:

(1) Demographic data: age, gender, type of school taught at, total years of teaching experience;
(2) Frequency of voice symptoms rated on a four-point rating scale (1 = no recurrence; 2 = monthly recurrence; 3 = weekly recurrence; 4 = daily recurrence);
(3) Severity of voice symptoms rated on a four-point rating scale (1 = none; 2 = mild; 3 = moderate; 4 = severe);
(4) Feeling of being in a noisy environment and having to raise their voices due to noise; both rated on a four-point rating scale (1 = always; 2 = sometimes; 3 = rarely; 4 = never);
(5) Location of schools (whether schools are located next to other schools, government offices and public sectors, and quiet streets with residential buildings or in open market areas);
(6) Location of classrooms (whether classrooms are located near or far from main traffic roads);
(7) Conditions of classrooms (approximate classroom area, number of students per class, number of windows in the classroom, the window and door material, any broken doors or windows, whether the windows and doors were closed during teaching, lighting/number of tube lights, the use of any ventilation, aeration, and insulation systems, and the presence of suspended ceilings based on a dichotomous response of yes/no).

Statistical analysis

Descriptive statistics were used to calculate percentages for the severity and frequency of voice symptoms and perception of noise. The percentage distribution was also calculated for other variables, namely noise sources, the location of schools and classrooms, and classroom conditions. The chi-squared test and Fisher's exact test (used when the chi-squared test assumption was violated by having more than 20% of cells with expected counts less than five) were used to find correlations between ordinal and nominal variables (ie, correlations between noise sources, the location of schools and classrooms, and the frequency and severity of voice symptoms). Goodman and Kruskal's Gamma was used to find correlations between ordinal variables (ie, correlations between the frequency and severity of voice symptoms and the perception of noise, voice raising, and classroom conditions). Additionally the Kruskal-Wallis test was used to compare the severity and frequency of voice symptoms for

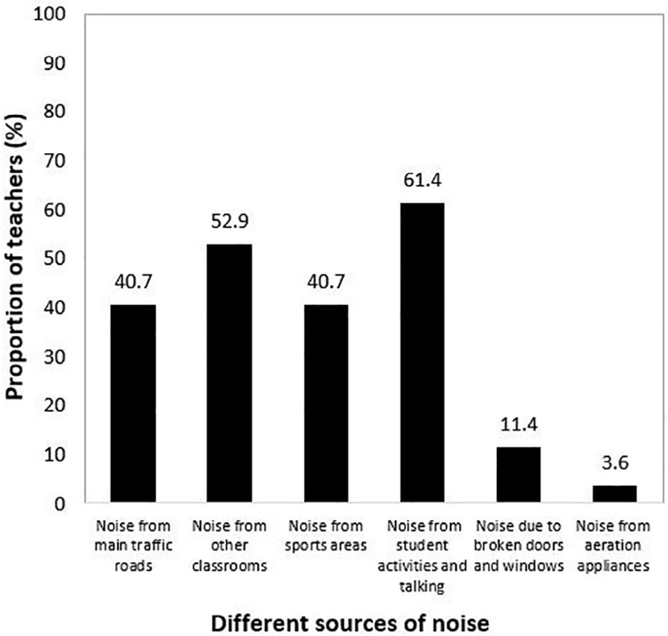


FIGURE 1. Percentage of teachers reporting noise from different sources.

teachers with less and more than 2 and 12 years of teaching experience. The significance level was set at $P < 0.05$ for all statistical analyses. The statistical analyses were carried out using *SPSS 22* software (IBM SPSS Statistics v. 22 for Windows, Armonk, NY).

Compliance with ethical standards

Ethical approval to carry out this study was obtained from Minia University's ethical committee and from undergraduate educational authorities in Minia, Egypt.

RESULTS

Relationship between teachers' self-reported voice symptoms, noise sources, noise perception, and number of years of teaching experience

The percentage of teachers reporting noise from different sources is shown in Figure 1. A total of 82.8% of teachers reported noise from more than one source. Noise from student activities and talking in their own classrooms was the most frequent source of noise (61.4% of teachers) followed by noise from other neighboring classrooms (52.9%). The chi-squared test showed a significant association between frequent laryngeal or neck pain symptoms and noise from other classrooms (χ^2(3, N = 140) = 18.786, $P < 0.001$). Of 44.8% of teachers who reported noise from neighboring classrooms, 13.5% reported experiencing daily recurrence and 9.4% experienced monthly recurrence of laryngeal pain (Table 1). The chi-squared test showed a significant association between increased voice effort to continue talking and noise from student activities and talking (χ^2(2, N = 140) = 7.281, $P = 0.026$). Of 61.4% of teachers who reported noise due to student activities and talking in their own class, 11.4% reported always requiring increased vocal effort to continue talking for long durations, while 27.1% reported the need for extra vocal effort only sometimes as shown in Table 2.

The results of the frequency distribution of teachers' reports of feeling they were in a noisy environment and having to raise

TABLE 1.
The Counts and Percentages of Teachers Reporting on the Frequency of Laryngeal or Neck Pain in Relationship to Noise From Neighboring Classrooms (N = 96 teachers)

	Noise From Neighboring Classroom		
Frequency of Laryngeal and Neck Pain	No	Yes	Total
No recurrence	11 (11.5%)	23 (24.0%)	34 (35.4%)
Monthly recurrence	14 (14.6%)	9 (9.4%)	23 (24%)
Weekly recurrence	18 (18.8%)	5 (5.2%)	23 (24%)
Daily recurrence	3 (3.1%)	13 (13.5%)	16 (16.7%)
Total	46 (47.9%)	50 (44.8%)	96 (100%)

TABLE 2.
The Counts and Percentages of Teachers Reporting on Increased Voice Effort in Relationship to Noise From Student Activities and Talking (N = 140 teachers)

	Noise From Student Activities and Talking		
Increased Voice Effort in Order to Continue Talking	No	Yes	Total
No voice effort	32 (22.9%)	32 (22.9%)	64 (45.7%)
Voice effort sometimes	13 (9.3%)	38 (27.1%)	51 (36.4%)
Voice effort always	9 (6.4%)	16 (11.4%)	25 (17.9%)
Total	54 (38.6%)	86 (61.4%)	140 (100%)

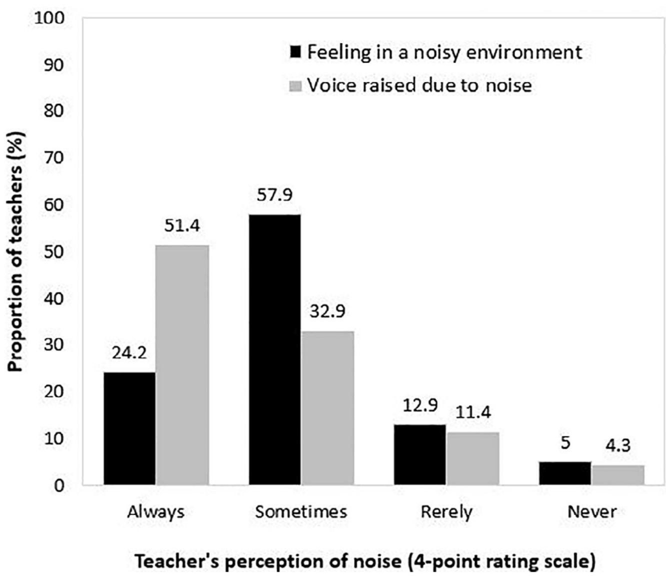


FIGURE 2. Frequency distribution of 140 teachers on their reports of feeling they were in a noisy environment and having to raise their voices due to noise.

their voices due to noise are shown in Figure 2. A total of 57.9% of teachers reported feeling they were sometimes in a noisy environment, and 24.2% reported always feeling they were in a noisy environment. On the other hand, 51.4% of teachers reported always having to raise their voice due to noise, while 32.9% reported having to raise their voice only sometimes. Goodman and Kruskal's Gamma showed a very strong positive association between feeling they were in a noisy environment and raising their voices due to it ($G = 0.876$, $P < 0.001$). In addition, correlations were found between raising one's voice and the severity and frequency of voice symptoms (Table 3).

Comparing the severity and frequency of voice symptoms for teachers with less and more than 2 and 12 years of teaching experience

The Kruskal-Wallis test was used to compare the severity and frequency of voice symptoms for teachers with less and more than 2 years of teaching experience and for those with less and more than 12 years of experience. Of 140 teachers, 22 (15.7%) had 2 and less years, while 118 (84.3%) teachers had more than 2 years of teaching experience. No significant differences ($P > 0.05$) in the voice symptom reports between these two groups were found. However, a significant difference was found on the frequency of dysphonia ($\chi^2(1) = 4.602$, $P = 0.03$), frequency of laryngeal or neck pain ($\chi^2(1) = 3.998$, $P = 0.04$), and severity of voice interrupted at the end of the day ($\chi^2(1) = 10.371$, $P = 0.001$) between teachers with 12 and less years of teaching experience (71 teachers [50.7%]) and teachers with more than 12 years of teaching experience (69 teachers [49.3%]). The mean ranks of the Kruskal-Wallis test were higher for teachers with more than 12 years of teaching experience.

Relationship between voice symptoms and school location

Of 140 teachers, 84 (60%) reported their school to be located close to other schools, 33 (23.6%) reported being close to other government offices, and 23 (16.4%) reported their schools to be situated close to quiet streets with residential buildings. None reported their schools to be located near market areas. The results of Fisher's exact test show a significant relationship between the severity of dysphonia and school location ($P = 0.012$). Table 4 shows that the teachers in schools close to other schools had more severe symptoms than teachers located close to quiet streets; these experienced only mild symptoms of dysphonia.

TABLE 3.
Goodman and Kruskal's Gamma Values and Significant *P* Values for Correlations Between Raising One's Voice Due to Noise and the Severity and Frequency of Voice Symptoms

Correlations Between		Type of Voice Symptom	Goodman and Kruskal's Gamma Value	Sig. (2-tailed) *P* value < 0.05
Raising voice due to noise	Severity of voice symptoms	Dysphonia	G = 0.327 (moderate)	*P* = 0.021
		Laryngeal or neck pain	G = 0.231 (weak)	*P* = 0.033
		Throat clearing	G = 0.298 (weak)	*P* = 0.007
		Interrupted voice at the end of the day with an inability to complete speech	G = 0.317 (moderate)	*P* = 0.031
	Frequency of voice symptoms	Throat clearing	G = 0.272 (weak)	*P* = 0.029
		Throat dryness	G = 0.239 (weak)	*P* = 0.042
	Increased voice effort in order to continue talking		G = 0.262 (weak)	*P* = 0.031

TABLE 4.
The Counts and Percentages of Teachers Reporting on the Severity of Dysphonia in Relationship to School Location Categories (N = 140 teachers)

Severity of Dysphonia	Location of Schools: Close to Other Schools	Close to Government Offices and Public Sectors	Close to Quiet Streets With Residential Buildings	Total
No dysphonia	19 (13.6%)	6 (4.3%)	1 (0.7%)	26 (18.6%)
Mild dysphonia	19 (13.6%)	14 (10%)	14 (10%)	47 (33.6%)
Moderate dysphonia	35 (25%)	11 (7.9%)	8 (5.7%)	54 (38.6%)
Severe dysphonia	11 (7.9%)	2 (1.4%)	0 (0%)	13 (9.3%)
Total	84 (60%)	33 (23.6%)	23 (16.4%)	140 (100%)

Relationship between voice symptoms and classroom location

A total of 55.2% of teachers reported their classroom to be close to main traffic roads, while 44.8% reported it to be far from main traffic roads. The chi-squared test showed a significant correlation between classroom location and the frequency of laryngeal and neck pain (χ^2[3, N = 96] = 9.48, $P = 0.02$). Here, however, reports of 44 teachers were missing, and only 96 teachers rated this variable, as shown in Table 5.

Relationship between voice symptoms and classroom conditions

Teachers reported an average classroom area of 27.3 m^2, with an average class size of 36 students. Results of Spearman rank-order correlations showed a significant correlation between the class size and the severity of throat dryness ($r = 0.18$, $P = 0.033$). The average number of windows per classroom was four with an average of eight tube lights per classroom. When asked about broken and fixed/unbroken doors and windows, 2 teachers reported broken windows, 56 (40%) reported broken doors, and 84 (60%) reported unbroken doors. All the teachers reported to have wooden doors in their classroom with 132 (94.3%) reporting windows made of only glass and 8 (6%) reporting windows made of both wood and glass. A weak but significant positive correlation was also found using Goodman and Kruskal's Gamma ($G = 0.257$, $P = 0.024$) between the frequency of dysphonia and the absence of closed doors and windows during teaching. From Table 6 it is clear that the largest number of teachers reported the doors and windows to be closed during teaching only sometimes. Of these 12 teachers (10.5%) had daily and 14 (12.3%) had monthly recurrence of dysphonia.

There were no suspended ceilings or sound insulators used in any of the classrooms (eg, insulated walls, window curtains,

TABLE 5.
The Counts and Percentages of Teachers Reporting on the Frequency of Laryngeal and Neck Pain in Relationship to Classroom Location (N = 96 teachers)

Frequency of Laryngeal and Neck Pain	Location of Classroom: Close to Main Traffic Roads	Far From Main Traffic Roads	Total
No recurrence	17 (17.7%)	17 (17.7%)	34 (35.4%)
Monthly recurrence	10 (10.4%)	13 (13.5%)	23 (24%)
Weekly recurrence	19 (19.8%)	4 (4.2%)	23 (24%)
Daily recurrence	7 (7.3%)	9 (9.4%)	16 (16.7%)
Total	53 (55.2%)	43 (44.8%)	96 (100%)

TABLE 6.
The Counts and Percentages of Teachers Reporting on the Frequency of Dysphonia in Relationship to Open/Closed Status of Doors and Windows During Teaching (N = 114 teachers)

Frequency of Dysphonia	Doors and Windows Closed During Teaching?: No	Rarely	Sometimes	Always	Total
No recurrence	6 (5.3%)	10 (8.8%)	24 (21.1%)	6 (5.3%)	46 (40.4%)
Monthly recurrence	3 (2.6%)	2 (1.8%)	14 (12.3%)	3 (2.6%)	22 (19.3%)
Weekly recurrence	3 (2.6%)	3 (2.6%)	9 (7.8%)	7 (6.1%)	22 (19.3%)
Daily recurrence	2 (1.8%)	2 (1.8%)	12 (10.5%)	8 (7%)	24 (21.1%)
Total	14 (12.3%)	17 (14.9%)	59 (51.8%)	24 (21.1%)	114 (100%)

TABLE 7.
Distribution of School Types by School Location Category

	Location of School			
Type of School	Close to Other Schools	Close to Government Buildings	Close to Quiet Streets With Residential Buildings	Total
Primary public	36 (25.7%)	0 (0%)	0 (0%)	36 (25.7%)
Primary private	6 (4.3%)	19 (13.6%)	8 (5.7%)	33 (23.6%)
Preparatory public	34 (24.3%)	0 (0%)	0 (0%)	34 (24.3%)
Preparatory private	8 (5.7%)	14 (10%)	15 (10.7%)	37 (26.4%)
Total	84 (60%)	33 (23.6%)	23 (16.4%)	140 (100%)

or carpeted floors). Regarding the use of ventilators and other aeration appliances, 129 (92%) teachers reported using fans, while 11 (8%) reported not using any type of ventilators or aeration appliances in their classrooms.

Relationship between school type and school location

There was a significant and strong correlation between school location and type of school (χ^2[6, N = 140] = 99.071, $P < 0.001$), with Cramer's V value being 0.595. Table 7 shows the distribution of school types by school location category as reported by 140 teachers. All public schools (N = 70) were situated close to other schools, while only 10% of private schools were close to other schools. Among the private schools, 16.4% were situated close to quiet streets with residential buildings.

DISCUSSION

This study aimed at identifying correlations between teachers' voice symptoms and their noise perception, noise resulting from the location of schools, and the location and conditions of classrooms. These sources of noise are often not investigated, particularly in terms of the negative impact they can have on a teacher's voice. We also sought to find out if years of teaching experience had an effect on the severity and frequency of voice symptoms.

Referring to section "Relationship between teachers' self-reported voice symptoms, noise sources, noise perception, and number of years of teaching experience" of the results, teachers reported noise to be sourced mainly from student activities and talking in their own classrooms (61.4% of teachers), as well as noise from neighboring classrooms (52.9%). This also had a negative effect on their frequent recurrence of laryngeal or neck pain (13.5% of teachers) (Table 1) and their need to exert extra vocal effort to continue talking for longer durations due to the noise in their own classrooms (11.4% of teachers reporting always and 27.1% reporting sometimes) (Table 2). In today's education system, students are encouraged to discuss and verbalize ideas and thoughts during active classroom situations, which may lead to increased levels of classroom noise.[36,38] The teachers then strain their voices and often speak aloud to discipline their students, which then leads to voice illnesses in these teachers, particularly due to vocal loading for long periods without any voice rests.[38] These voice problems may become aggravated when there is also noise from adjacent classrooms adding to the total noise, particularly when there are thin walls or partitions (made of poor acoustic insulation material) separating the classrooms. This leads to poor classroom acoustical quality, which is a common problem in Egyptian schools.[39,40]

Teachers also reported feeling they were in a noisy environment (57.9% of teachers reported sometimes and 24.2% reported always, from Figure 2) and had to raise their voices due to this noise (51.4% of teachers reported always and 32.9% reported sometimes, from Figure 2), which is related to a well-known phenomenon called the "Lombard effect."[41] This is a reflexive behavior of the speaker who involuntarily increases his/her vocal intensity and fundamental frequency due to noise, thus coping with the constraints posed by the noise on the reception of acoustic signal (in this case, the speaker's own voice). Due to the increase in vocal loudness and pitch, there is increased vocal effort,[42] vocal loading, vocal fatigue, and teacher stress.[18] In a study by Kristiansen et al,[6] authors reported a 0.65 dB(A) increase in vocal load with per dB(A) increase in noise level in classrooms, with teachers raising their voices 61% of the time. Speaking with a raised voice and increased vocal load seemed to cause dysphonia, laryngeal and neck pain, and throat dryness with frequent throat clearing among the teachers. These teachers also reported not being able to maintain their speaking at the end of the day and exerting much vocal effort to continue talking. These findings are evident from Table 3, where significant, weak to moderate correlations were obtained between the raising of voice due to noise and the severity and frequency of different voice symptoms.

Considering the section "Comparing the severity and frequency of voice symptoms for teachers with less and more than 2 and 12 years of teaching experience," we compared voice symptoms in teachers with different years (less and more than 2 and 12 years) of teaching experience. From the results of the previous study by Abo-Hasseba et al[29] on the same study population, it was found that teachers working in the public schools had an average of 17.9 years of teaching experience, while those working in private schools had an average of 7.4 years of teaching experience with a total average of 12.3 years of teaching experience combining both school types. They also reported that no significant differences were found between public and private teachers for the severity and frequency of voice symptoms. In the present study as well, no significant difference in the severity and frequency of voice symptoms was found between groups with less and more than 2 years of teaching experience, probably due to the smaller sample size of the group with teaching experience of 2 years and less (only 15.7% teachers). However, considering the total average of 12.3 years of teaching experience combining both school types,[29] we sought to compare the severity and frequency of voice symptoms for the teachers with 12 and less versus more than 12 years of teaching experience. It was found that teachers with more than 12 years of teaching experience had significantly higher scores on frequency of dysphonia, laryngeal or neck pain, and severity of voice interrupted at the end of the working day. The more the years of teaching, the more the teachers are exposed to unfavorable environmental conditions, particularly teaching in a noisy classroom for many years without any appropriate voice use and vocal hygiene. Literature also proves that a long job tenure (longer years of teaching) may be a crucial factor to the development of long-standing pharyngeal and laryngeal abnormality and hyperfunction signs in teachers due to long-term exposure to noise.[43,44]

Considering school site (the section "Relationship between voice symptoms and school location"), a school that is located in a busy and noisy area (eg, near traffic or an industrial area) is highly unfavorable for both learning and working conditions. The World Health Organization, in their document titled "Information series of school health-Document 2," define a health-promoting school as "one that constantly strengthens its capacity as a healthy setting for living, learning and working,"[45] This document provides guidelines for selecting a school site by considering the potential environmental risk in terms of a school's location, including the risk of noise. In the present study (Table 4), about 60% of teachers reported their schools were located close to other schools. Of these teachers, 13.6% reported having mild, 25% reported having moderate, and 7.9% reported having severe dysphonia. A total of 23.6% of teachers had schools located close to other government buildings (such as government banks, passport offices, and directorate service offices, which are very crowded and noisy). A significant association was found between the school location and the severity of dysphonia in these teachers. When two or more schools or other noisy government offices are situated close to each other, there is an additive effect on the total noise generated from nearby schools and public sectors. Only 16.4% of teachers reported their schools to be situated close to quiet streets with residential buildings, which may be a more favorable environment. However, even these teachers reported having mild to moderate dysphonia, which may be due to unawareness or poor vocal hygiene and improper use of voice. Another factor may be teaching without breaks (no voice rest), as seen in 61% of teachers in the previous study[29] conducted on the same study population. When schools are located in a quiet residential area with more trees and plants around, the trees are able to absorb some of the sound, thus reducing noise levels. The acoustic benefits of trees and plants in attenuating noise have been very well known for many years.[46] Currently, there have been actions taken to promote "Green Schools" by the U.S. Green Building Council[47] in order to construct school buildings that provide a healthy environment for the improved performance of teachers and students. In Egypt, schools are mostly located adjacent to other schools or in noisy environments. In a most recent Egyptian study,[48] authors provide guidelines and selection procedures for the appropriate planning of school sites, taking into consideration geographical, environmental, educational, technical, and safety aspects.

When correlating voice symptoms and classroom location and conditions (the sections "Relationship between voice symptoms and classroom location" and "Relationship between voice symptoms and classroom location"), all teachers who reported their classrooms to be close to main traffic roads (55.2%) experienced daily (7.3%), weekly (19.8%), and monthly (10.4%) recurring frequency of laryngeal or neck pain (Table 5). Voice symptoms such as laryngeal pain, throat clearing, and throat dryness may be attributed to excess vocal loading, vocal fatigue, and speaking in a raised voice (or increased loudness) for continuous long periods.[49–51] The possible reasons for excess classroom noise due to road traffic could be due to the finding that only 51.8% of teachers reported doors and windows in their classroom to be closed only sometimes while teaching, while in 14.9% of teachers reported that they were closed rarely and in 12.3% they were never closed (Table 6); additionally, 40% of teachers reported broken doors, which might have aggravated the noise levels in classrooms. Sound insulation windows with monolayer or two double-hollow glass layers should be used, as they attenuate external noise (road traffic noise) up to 32 and 24 dBSPL, respectively.[52] Such considerations to school classrooms may help teachers to preserve good vocal health, as they reduce vocal loading resulting from noise. Road traffic noise is the primary type of troublesome external noise.[9]

About 92% of teachers reported using fans as a ventilation system. There was also an average of eight tube lights inside every classroom. Noise from ventilation systems (fans in this study) and power lines are all sources of low-frequency noise. The low-frequency noise (20–200 Hz) usually emitted by heating-ventilation and air-conditioning systems occurs due to less attenuation of these unwanted sounds by walls and floors.[53,54] The presence of this low-frequency noise is a reason for the reported unpleasantness and annoyance of noise in humans.[55,56] For reducing low-frequency internal noise, the installation of thick cotton curtains and porous floor carpets, which may decrease up to 2.4–4.5 dB of internal noise, should be considered.[11,57]

Communication through spoken language is essential for students learning in an active classroom, but when there are 30–40 students (average of 36 students in this study) accommodated in one class, their loud verbalizations may increase noise levels in the classrooms. Overcrowded classrooms are a reason for elevated stress in teachers, who find it very difficult to maintain control and teach efficiently. This may cause teachers to use their voices in a strained manner, affecting their vocal health. In a study conducted by Munier and Farrell,[58] they reported teachers with a class size greater than 30 to have more frequent voice problems than those teaching fewer students. A similar result was seen in a study where teachers of larger class sizes had a threefold increased risk of having voice symptoms.[59] Due to the increased Egyptian population and the mandatory rule of the state to provide education to all children 6 years and older, schools enroll students into classrooms of more than 40–50 students,[28,60,61] the implications of which require further consideration, including the need for class size standards in Egyptian schools.

Lastly we also sought to identify associations between school type and the location of schools. We found that all public schools (N = 70, which are government funded) are situated closer to other schools, as compared with private schools, which are located closer to other government buildings (23.6%) and quiet streets with residential buildings (16.4%, Table 7). Also, it was found here that teachers working in public schools had more voice and throat symptoms and more often felt they were in a noisy environment than teachers working in private schools, as seen in the previous study conducted by Abo-Hasseba et al[29] on the same study population. This study also mentions that public schools are overcrowded when compared with private schools, which may be one reason for the poor voice quality in public school teachers. Moreover, the public schools located close to other schools may have an additive external noise effect in the public school classrooms, causing these teachers to have a higher degree of voice problems.

The present study advocates considering noise from inappropriate school and classroom locations and poor classroom conditions to be causative factors for the prevalence of voice symptoms in teachers of Egyptian schools. This study, which is based on teacher self-reports, encourages future studies measuring internal and external noise levels of Egyptian schools to consider these factors. This would further bring about useful applications for the improved planning of new Egyptian schools and/or the renovation of existing schools. This improved planning may be in terms of location and infrastructure for improved classroom conditions with optimum acoustical quality similar to those recommended by standards in the United States.[11,62] Such an initiative would be conducive for both teachers' vocal health and a better learning place for students.

CONCLUSION

This study showed that noise resulting from inappropriate school and classroom locations and poor classroom conditions had moderate to severe repercussions on a teacher's voice. This requires attention and necessitates solutions for the future improvement of Egyptian schools and classrooms. Significant correlations between voice symptoms and inappropriate environmental conditions suggest that teachers' vocal conditions are very vulnerable to undesirable environments, including factors not typically considered to have an effect. Teachers reported feeling they were in a noisy environment most of the time and had to raise their voice due to this noise, thus straining their vocal organ. This study forms a base for future studies that could be conducted in Egyptian schools, focusing on recommending appropriate architecture and infrastructure standards that consider the health of teachers and students. The present study suggests the need for designing schools in a better way so as to achieve a positive and successful work/learning environment. Also, it suggests that teachers should be provided with vocal training programs to overcome undesirable environmental conditions and use their voices appropriately in this vocally demanding occupation.

Acknowledgments

The study was accomplished during the research stay of Ketaki Vasant Phadke at the University of Helsinki. Her stay was supported by the funds from Erasmus Plus studies within program countries (No. 2017–2018/164).

REFERENCES

1. Levak K, Horvat M, Domitrovic H. Effects of Noise on Humans. ELMAR, 2008. 50th International Symposium. Vol 1: IEEE; 2008:333–336.
2. Guski R, Felscher-Suhr U, Schuemer R. The concept of noise annoyance: how international experts see it. *J Sound Vib* 1999;223:513–527. doi:10.1006/jsvi.1998.2173.
3. Cutiva LCC, Puglisi GE, Astolfi A, et al. Four-day follow-up study on the self-reported voice condition and noise condition of teachers: relationship between vocal parameters and classroom acoustics. *J Voice*. 2017;31:120.e1–120.e8. doi:10.1016/j.jvoice.2016.02.017.
4. Durup N, Shield BM, Dance S, et al. Teachers' voice parameters and classroom acoustics—a field study and online survey. *J Acoust Soc Am*. 2017;141:3540. doi:10.1121/1.4987482.
5. Rantala LM, Hakala S, Holmqvist S, et al. Classroom noise and teachers' voice production. *J Speech Lang Hear Res*. 2015;58:1397–1406. doi:10.1044/2015_JSLHR-S-14-0248.
6. Kristiansen J, Lund SP, Persson R, et al. A study of classroom acoustics and school teachers' noise exposure, voice load and speaking time during teaching, and the effects on vocal and mental fatigue development. *Int Arch Occup Environ Health*. 2014;87:851–860. doi:10.1007/s00420-014-0927-8.
7. Ilomäki I, Leppänen K, Kleemola L, et al. Relationships between self-evaluations of voice and working conditions, background factors, and phoniatric findings in female teachers. *Logoped Phoniatr Vocol*. 2009;34:20–31. doi:10.1080/14015430802042013.
8. Thibeault SL, Merrill RM, Roy N, et al. Occupational risk factors associated with voice disorders among teachers. *Ann Epidemiol*. 2004;14:786–792. doi:10.1016/j.annepidem.2004.03.004.
9. Shield B, Dockrell JE. External and internal noise surveys of London primary schools. *J Acoust Soc Am*. 2004;115:730–738. doi:10.1121/1.1635837.
10. Harght P, Coffeen RC. Comparing classroom acoustics in green and nongreen schools. *J Acoust Soc Am*. 2008;124:2588. doi:10.1121/1.4783209.
11. ANSI.S12.60-2002. American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools. Melville, NY: Acoustical Society of America; 2002:1–36.
12. Koszarny Z, Goryński P. Exposure of schoolchildren and teachers to noise at school. *Rocz Panstw Zakl Hig*. 1990;41:297–310.
13. Rogerson J, Dodd B. Is there an effect of dysphonic teachers' voices on children's processing of spoken language? *J Voice*. 2005;19:47–60. doi:10.1016/j.jvoice.2004.02.007.
14. Nelson P. The changing demand for improved acoustics in our schools. *Volta Rev*. 1999;101:23–31.
15. Åhlander VL, Rydell R, Löfqvist A. Speaker's comfort in teaching environments: voice problems in Swedish teaching staff. *J Voice*. 2011;25:430–440. doi:10.1016/j.jvoice.2009.12.006.
16. Vilkman E. Voice problems at work: a challenge for occupational safety and health arrangement. *Folia Phoniatr Logop*. 2000;52:120–125. doi:10.1159/000021519.
17. Cutiva LCC, Vogel I, Burdorf A. Voice disorders in teachers and their associations with work-related factors: a systematic review. *J Commun Disord*. 2013;46:143–155. doi:10.1016/j.jcomdis.2013.01.001.
18. Tiesler G, Oberdörster M. Noise—a stressor? Acoustic ergonomics of schools. *Build Acoust*. 2008;15:249–261.
19. Berg FS, Blair JC, Benson PV. Classroom acoustics: the problem, impact, and solution. *Lang Speech Hear Serv Sch*. 1996;27:16–20. doi:10.1044/0161-1461.2701.16.
20. Bottalico P, Astolfi A, Hunter EJ. Teachers' voicing and silence periods during continuous speech in classrooms with different reverberation times. *J Acoust Soc Am*. 2017;141:EL26–EL31. doi:10.1121/1.4973312.
21. Puglisi GE, Astolfi A, Cantor Cutiva LC, et al. Four-day-follow-up study on the voice monitoring of primary school teachers: relationships with

conversational task and classroom acoustics. *J Acoust Soc Am*. 2017;141:441–452. doi:10.1121/1.4973805.
22. Astolfi A, Carullo A, Pavese L, et al. Duration of voicing and silence periods of continuous speech in different acoustic environments. *J Acoust Soc Am*. 2015;137:565–579. doi:10.1121/1.4906259.
23. Durup N, Shield B, Dance S, et al. An investigation into relationships between classroom acoustic measurements and voice parameters of teachers. *Build Acoust*. 2015;22:225–241.
24. Hunter EJ, Bottalico P, Graetzer S, et al. Teachers and teaching: speech production accommodations due to changes in the acoustic environment. *Energy Procedia*. 2015;78:3102–3107. doi:10.1016/j.egypro.2015.11.764.
25. Åhlander VL, García DP, Whitling S, et al. Teachers' voice use in teaching environments: a field study using ambulatory phonation monitor. *J Voice*. 2014;28:841.e5–841.e15. doi:10.1016/j.jvoice.2014.03.006.
26. Bottalico P, Astolfi A. Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *J Acoust Soc Am*. 2012;131:2817–2827. doi:10.1121/1.3689549.
27. Pelegrín-García D, Brunskog J. Speakers' comfort and voice level variation in classrooms: laboratory research. *J Acoust Soc Am*. 2012;132:249–260. doi:10.1121/1.4728212.
28. Bolbol SA, Zalat MM, Hammam RA, et al. Risk factors of voice disorders and impact of vocal hygiene awareness program among teachers in public schools in Egypt. *J Voice*. 2017;31:251.e9–251.e16. doi:10.1016/j.jvoice.2016.07.010.
29. Abo-Hasseba A, Waaramaa T, Alku P, et al. Difference in voice problems and noise reports between teachers of public and private schools in Upper Egypt. *J Voice*. 2017;31:508.e11–508.e16. doi:10.1016/j.jvoice.2016.10.016.
30. AboulKheir AE. Consideration of primary school design standards with reference to Egyptian schools [doctoral dissertation]; 1967. doi:10.3929/ethz-a-000089241.
31. Ibrahim N. Efforts for improving buildings conditions in public schools in Egypt: SWOT analysis of two governmental renovation programs. [master thesis]. American University in Cairo; 2017. Available at http://dar.aucegypt.edu/handle/10526/4977. Accessed March 27, 2018.
32. El Baradei M, El Baradei L. Needs assessment of the education sector in Egypt. ZEF Bildungsstudie, Cairo; 2004. Available at https://www.zef.de/fileadmin/webfiles/downloads/projects/el-mikawy/egypt_final_en.pdf. Accessed March 27, 2018.
33. Hamdy A. Survey of ICT and education in Africa: Egypt Country Report; 2007. Available at http://www.infodev.org/infodev-files/resource/InfodevDocuments_399.pdf. Accessed March 27, 2018.
34. Ali SA. Investigation of the dose–response relationship for road traffic noise in Assiut, Egypt. *Appl Acoust*. 2004;65:1113–1120. doi:10.1016/j.apacoust.2004.06.007.
35. Ali SA. Railway noise levels, annoyance and countermeasures in Assiut, Egypt. *Appl Acoust*. 2005;66:105–113. doi:10.1016/j.apacoust.2004.06.005.
36. Ali SA. Study effects of school noise on learning achievement and annoyance in Assiut City, Egypt. *Appl Acoust*. 2013;74:602–606. doi:10.1016/j.apacoust.2012.10.011.
37. Zaki GR. Assessment of ambient noise levels in the urban residential streets of Eastern Alexandria, Egypt. *J Egypt Public Health Assoc*. 2012;87:96–103. doi:10.1097/01.EPX.0000421367.39770.52.
38. Martin S, Darnley L. *The Teaching Voice*. 2 ed. London: Whurr; 2004.
39. Awad HS, Farag HH, Hanafi MA, et al. Architectural acoustics in educational facilities: An empirical study on university classrooms in Egypt. Proceedings of Meetings on Acoustics 164ASA. Vol 18: ASA; 2012:015002. doi:10.1121/1.4772593.
40. Herrera L. Participation in school upgrading: gender, class and (in) action in Egypt. *Int J Educ Dev*. 2003;23:187–199. doi:10.1016/S0738-0593(02)00013-5.
41. Lombard E. Le signe de I'elevation de la voix, Annals Maladiers Oreille. *Larynx Nez Pharynx*. 1911;37:101–119.
42. Lee J, Ali H, Ziaei A, et al. The Lombard effect observed in speech produced by cochlear implant users in noisy environments: a naturalistic study. *J Acoust Soc Am*. 2017;141:2788–2799. doi:10.1121/1.4979927.
43. Augustyńska D, Kaczmarska A, Mikulski W, et al. Assessment of teachers' exposure to noise in selected primary schools. *Arch Acoust*. 2010;35:521–542. doi:10.2478/v10168-010-0040-2.
44. Bronder A. *Study of phenomenon of voice disorders in the population of teachers and the prevention rules*, [doctor's thesis]. Institute of Occupational Medicine and Environmental Health in Sosnowiec, Sosnowiec, Poland; 2003.
45. World Health Organization. The physical school environment: an essential element of a health-promoting school; 2004. Available at http://www.who.int/iris/handle/10665/42683. Accessed March 27, 2018.
46. Van Renterghem T, Botteldooren D, Verheyen K. Road traffic noise shielding by vegetation belts of limited depth. *J Sound Vib* 2012;331:2404–2425. doi:10.1016/j.jsv.2012.01.006.
47. Gordon DE. Green schools as high performance learning facilities. National Clearinghouse for Educational Facilities; 2010. Available at http://files.eric.ed.gov/fulltext/ED512700.pdf. Accessed March 27, 2018.
48. Moussa M, Mostafa Y, Elwafa A. School site selection process. *Procedia Environ Sci*. 2017;37:282–293. doi:10.1016/j.proenv.2017.03.059.
49. Vintturi J, Alku P, Sala E, et al. Loading-related subjective symptoms during a vocal loading test with special reference to gender and some ergonomic factors. *Folia Phoniatr Logop*. 2003;55:55–69. doi:10.1159/000070088.
50. Laukkanen A-M, Järvinen K, Artkoski M, et al. Changes in voice and subjective sensations during a 45-min vocal loading test in female subjects with vocal training. *Folia Phoniatr Logop*. 2004;56:335–346. doi:10.1159/000081081.
51. Laukkanen A-M, Kankare E. Vocal loading-related changes in male teachers' voices investigated before and after a working day. *Folia Phoniatr Logop*. 2006;58:229–239. doi:10.1159/000093180.
52. Zhai G, Zhang B. The design of ventilation and sound insulation window. *Noise Vib Control*. 2004;1:014.
53. Berglund B, Hassmen P, Job RS. Sources and effects of low-frequency noise. *J Acoust Soc Am*. 1996;99:2985–3002. doi:10.1121/1.414863.
54. Waye KP, Rylander R, Benton S, et al. Effects on performance and work quality due to low frequency ventilation noise. *J Sound Vib* 1997;205:467–474. doi:10.1006/jsvi.1997.1013.
55. Leventhall G, Pelmear P, Benton S. A review of published research on low frequency noise and its effects. 2003. Available at http://westminsterresearch.wmin.ac.uk/4141/1/Benton_2003.pdf. Accessed March 27, 2018.
56. Baliatsas C, van Kamp I, van Poll R, et al. Health effects from low-frequency noise and infrasound in the general population: is it time to listen? A systematic review of observational studies. *Sci Total Environ*. 2016;557:163–169. doi:10.1016/j.scitotenv.2016.03.065.
57. Zhisheng L, Dongmei L, Sheng M, et al. Noise impact and improvement on indoors acoustic comfort for the building adjacent to heavy traffic road. *Chin J Popul Resour Environ*. 2007;5:17–25. doi:10.1080/10042857.2007.10677482.
58. Munier C, Farrell R. Working conditions and workplace barriers to vocal health in primary school teachers. *J Voice*. 2016;30:127.e31–127.e41. doi:10.1016/j.jvoice.2015.03.004.
59. Thomas G, Kooijman P, Cremers C, et al. A comparative study of voice complaints and risk factors for voice complaints in female student teachers and practicing teachers early in their career. *Eur Arch Otorhinolaryngol Head Neck*. 2006;263:370–380. doi:10.1007/s00405-005-1010-6.
60. Egypt; World data on education; 2012. Available at http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/Egypt.pdf. Accessed March 27, 2018.
61. Ministry of Education. National strategic plan for pre-university education reform in Egypt 2007/08–2011/12. Cairo; 2007.
62. American National Standards Institute/Acoustical Society of America. S12.60-2009/Part 2 American National Standard Acoustical Performance Criteria DR, and Guidelines for Schools, Part 2: Relocatable Classroom Factors. Melville, NY: Acoustical Society of America; 2009.