Department of Biophysics
Faculty of Science
Palacký University Olomouc

Investigation of glottal configurations in singing

Doctoral Dissertation, December 2011

Christian T. HERBST

Supervisor:

RNDr. Jan G. Švec, Ph.D. et Ph.D.

Herbst, Christian T.

Investigation of glottal configurations in singing

[Výzkum nastavování hlasivkové štěrbiny při zpěvu]

Doctoral Thesis, Palacký University in Olomouc, the Czech Republic - with an abstract in Czech.

Table of Contents

Acknowledgements

Personal thanks

It is customary to thank one's supervisor at this point in the dissertation. I am, however, happy to say that I express my sincere thanks to Jan Švec not out of custom or obligation. I had the unique chance of working with and learning from an incredibly smart and knowledgeable researcher. He was dedicated and precise when needed, and yet very supportive and friendly at the same time. I want to express my deep gratitude to Jan Švec for all the hours he invested, and I readily concede that without his help and support my scientific work would have never reached its current level.

In the recent years, I have had the chance to meet several outstanding persons that had a great influence on me: I want to thank W. Tecumseh Fitch for his enthusiastic encouragement, and for the unique opportunity to work in his lab. My heartfelt thanks go to David Howard, Sten Ternström and Donald Miller, for supporting me in the years before I started my dissertation. Many thanks to Michael Edwards for his friendship and guidance during my first attempts at sound analysis and digital signal processing. Finally, I want to thank Josef Schlömicher-Thier for continuing support, and for dragging me to the AQL 2001 in Groningen, where I became irrevocably addicted to voice science.

I would not have been able to finish my doctoral studies and this dissertation without the never ceasing love and support from my wife Angelika, who is the mother of three of my four wonderful children. Embarking on the PhD studies was a commitment for her as it was for me. The least I can do as a token of my gratitude is to dedicate this dissertation to her.

Electroglottographic Wavegrams

This research was supported by the Grant Agency of the Czech Republic, project GAČR 101/08/1155, by the ERC SOMACCA grant, and by the University of Vienna.

Four phonation types – pilot study

The data for this study were obtained in 2004 during the authors' stays at the Department of Speech, Music and Hearing, KTH, Stockholm. C. Herbst's stay was supported by the Erasmus Student Exchange Programme of the European Commission, and J. Švec's stay was supported by an individual grant from the Wenner-Gren Foundation. In 2008, the study was supported by the Grant Agency of the Czech Republic, project GAČR 101/08/1155. My thanks go to Hans Larsson at the Department of Logopedics and Phoniatrics, Karolinska University Hospital at Huddinge, Stockholm, for his help in acquiring the laryngoscopic recordings.

In the Netherlands, the research was supported by the Technology Foundation STW (Stichting Technische Wetenschappen) project GKG5973, Applied Science Division of NWO (Natuurwetenschappelijk Onderzoek), and the technology program of the Ministry of Economic Affairs, the Netherlands. In the Czech Republic, the work was supported by the Grant Agency of the Czech Republic, project GAČR 101/08/1155. I kindly thank Dr. D. Lazár, from the Department of Biophysics, Palacký University Olomouc, for his help with the statistic analysis of the data.

My sincere thanks go to my colleague Hana Šrámková, for her help in printing this dissertation.

Statement of originality

I declare that I am the sole author of this thesis and that all sources are properly cited.

Investigation of glottal configurations in singing

Abstract

The source of the human voice originates in the larynx and is generated by the vibrating vocal folds. This dissertation provides new insights into the voice source in singing in a twofold way:

The first aim of this dissertation was to develop a **new method for visualization and analysis of the electroglottographic (EGG) signal** (i.e. a physiological correlate of vocal fold vibration). This new method, termed "**EGG wavegram**", allows to display EGG signals across various phonations in one graph, whilst retaining the original appearance of the EGG waveform. The EGG signal is decomposed into consecutive individual cycles, each of which is normalized in both duration and amplitude, and is displayed on the y-axis, going from bottom to top. Overall time is shown on the x-axis. In a **DEGG wavegram**, the first derivative of the EGG signal is used as the input signal. In such a display, the contacting and de-contacting phases for each glottal cycle are approximated by (a) one or more dark horizontal line(s) at the lower end of the graph (contacting phase), and (b) one or more light horizontal line(s) in the upper section of the graph (de-contacting phase). Much like in the sound spectrogram, information on vibratory behavior developing in time is compacted into one single graph, providing insight into changes of vocal fold dynamics. As such, the wavegram allows intuitive assessment of the time-varying contact phase of phonation over a longer period of time, indicating physiological changes of laryngeal configuration, such as vocal register. EGG wavegrams promise to be useful in research, clinical diagnostics, voice therapy and voice pedagogy.

The second aim of the dissertation was to gain more understanding of the laryngeal adjustments when producing various singing voice qualities. Two **basic adjustments of glottal configuration** are found and described: (a) **Cartilaginous adduction**, i.e. adduction of the posterior glottis. This is controlled by the singer when changing the quality of voice between "breathy" and "pressed". Phonation with a fully adducted glottis is characterized by strong high-frequency components (partials), thus giving the voice a "brassy" quality. Phonation with a posterior glottal gap creates high-frequency partials of lesser strength, resulting in a "flutey" or "dull" quality, and possibly creates noise components ("breathy voice"). (b) **Membranous medialization**, i.e. medial bulging of the vocal folds. This is controlled by adjusting the singing voice quality between chest and falsetto registers. In chest register, the acoustic spectrum of voice contains strong high-frequency partials. Phonation in falsetto register, on the other hand, usually results in a less "resonant" sound, containing weaker overtones.

The gathered evidence suggests that these two maneuvers can be controlled separately by both trained and untrained singers. Such a finding has not been documented in the scientific literature to date. The ability to individually and gradually control cartilaginous adduction and membranous medialization allows experienced singers to produce a variety of vocal timbres at the laryngeal level, thus increasing the quality of their artistic performance. This concept is also very promising for voice pedagogy and therapy, and for better understanding various singing styles.

Výzkum nastavování hlasivkové štěrbiny při zpěvu

Souhrn

Zdroj hlasu v hrtanu je tvořen kmitajícími hlasivkami. Tato dizertační práce přináší nové poznatky o zdroji hlasu při zpěvu, a to ze dvou pohledů: (a) představuje novou metodu "EGG wavegramů", která zviditelňuje a analyzuje elektroglotografický (EGG) signál, který je fyziologickým korelátem kmitání hlasivek; a (b) studuje fyziologická nastavení hrtanu při tvorbě různých kvalit zpěvního hlasu.

První cílem práce bylo vyvinout novou metodu pro vizualizaci a analýzu EGG signálu, která umožní zobrazení EGG signálu během různých fonací v jednom grafu, přičemž zachová informaci o průběhu individuálních cyklů kmitů hlasivek. Pro tento účel byla vyvinuta metoda tzv. „EGG wavegramů". EGG signál je zde rozdělen na jednotlivé postupné cykly kmitů, každý z nich je normalizován vzhledem k trvání periody a amplitudě signálu a poté je individuálně zobrazen podél osy y ve směru zdola nahoru. Celkový čas je zobrazen na ose x. U "DEGG wavegramu" je jako vstupní signál použita první derivace EGG signálu. V tomto zobrazení jsou fáze zvětšujícího a zmenšujícího se kontaktu hlasivek zviditelněny ve formě (a) jedné nebo více horizontálních čar ve spodní části grafu (fáze zvětšujícího se kontaktu) a (b) jedné nebo více horizontálních čar v horní části grafu (fáze zmenšujícího se kontaktu). Podobně jako u zvukového spektrogramu je informace o změnách kmitání hlasivek v čase komprimována do jediného obrázku, který tak poskytuje komplexní vhled do změn dynamiky kmitání hlasivek. Wavegram umožňuje intuitivní hodnocení periodických změn kontaktu mezi hlasivkami v delších časových usecích a napovídá o fyziologických změnách nastavení hlasivek, včetně hlasových rejstříků. EGG wavegram se jeví jako nadějná metoda jak pro výzkum, tak pro hlasovou diagnostiku, terapii a pedagogiku.

Druhým cílem práce bylo lepší porozumění možností nastavení hrtanu při tvorbě různých kvalit zpěvního hlasu. Jsou zde identifikovány a popsány **dva základní druhy nastavení hlasivkové štěrbiny**: (a) **Addukce chrupavčité glottis**, t.j. zadní části hlasivkové štěrbiny. Tato addukce je zpěvákem regulována při změnách mezi „dyšnou" a „tlačenou" kvalitou hlasu. Fonace s plně uzavřenou (addukovanou) glottis je charakteristická silnými harmonickými tóny, které propůjčují hlasu „ostřejší" (angl. „brassy") kvalitu. Fonace při otevřené (abdukované) chrupavčité části glottis je charakteristická spíše slabšími vyššími harmonickými tóny, což vede k spíše chudší „flétnové" (angl. „flutey" či „dull") kvalitě hlasu a k možnému výskytu šumových složek spektra ("dyšný hlas"). (b) **Medializace blanité části glottis**, t.j. vyklenutí mediální části hlasivek je modifikováno při změnách hlasu mezi hrudním a falzetovým rejstříkem. V hrudním rejstříku obsahuje akustické spektrum hlasu relativně intenzivní harmonické tóny. Na druhou stranu falzetový rejstřík vede spíše k „méně rezonančnímu" hlasu obsahujícímu harmonické tóny o nižší intenzitě.

Shromážděná evidence vypovídá o tom, že tyto dva druhy nastavení hlasivkové štěrbiny mohou být ovládány nezávisle jak u školených tak u neškolených zpěváků. Tento jev doposud nebyl v odborné literatuře popsán. Schopnost individuálně a spojitě ovládat addukci chrupavčité glottis a medializaci blanité části glottis umožňuje zpěvákům vytvářet různou barvu hlasu již na úrovni hrtanu, a tím zvyšovat uměleckou kvalitu jejich projevu. Tento koncept se ukazuje jako nosný i pro hlasovou pedagogiku a pro lepší pochopení mechanismu tvorby různých stylů zpěvního projevu.

List of publications in this thesis

The thesis is based on the following publications, which are referred to in the text according to their bold roman numerals:

I. Herbst, C. T., Fitch, W. T., and Švec, J. G. (**2010**). "Electroglottographic wavegrams: a technique for visualizing vocal fold dynamics noninvasively," Journal of the Acoustical Society of America **128**, 3070-3078.

cited in: Kob et al., 2011; Unger et al., 2011; Ziethe et al., 2011

II. Herbst, C. T., Ternström, S., and Švec, J. G. (**2009**). "Investigation of four distinct glottal configurations in classical singing - a pilot study," Journal of the Acoustical Society of America **125**, EL104-EL109.

cited in: Echternach et al., 2011b; Kochis-Jennings et al., in press; McCulloch et al., 2011

III. Herbst, C. T., Qiu, Q., Schutte, H. K., and Švec, J. G. (**2011**). "Membranous and cartilaginous vocal fold adduction in singing," Journal of the Acoustical Society of America **129**, 2253-2262.

cited in: Echternach et al., 2011a

I hereby confirm that I am the main author of these publications.

1. Introduction

The topic of this interdisciplinary dissertation is of interest for practitioners of both science and art. In order to reach such a broad audience, basic principles of voice production and measurement are reviewed here.

1.1. Basic functional principle of voice production

Voice is – with a few exceptions - created by vibrating laryngeal tissue (i.e. the vocal folds), which converts a steady airflow, as supplied by the lungs, into a sequence of flow pulses. The acoustic pressure waveform resulting from this sequence of flow pulses excites the vocal tract, which filters them acoustically, and the result is radiated from the mouth and sometimes from the nose. This description of the voice production mechanism suggests three basic physiologic layers for sound quality modification: the respiratory system, the larynx, and the vocal tract. These three sub-systems are sometimes termed the *power source*, *sound source* and *sound modifiers* (Howard and Murphy, 2007) - see Figure 1 for a rough anatomical (left) and a schematic (right) representation of the three sub-systems.

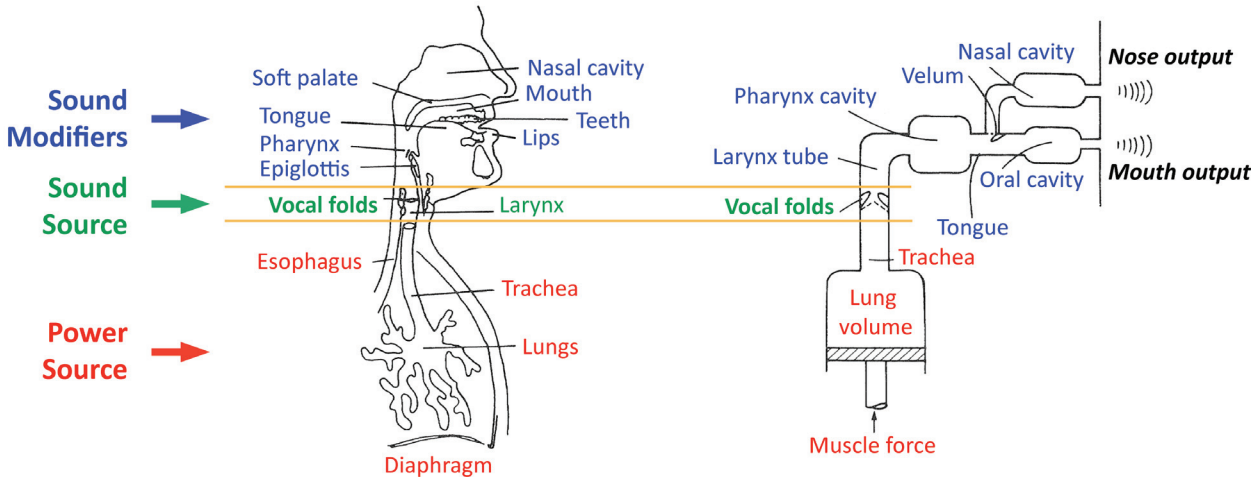


Figure 1 - Human vocal organs and a representation of their main acoustical features – Rossing, (1990), Fig. 15.1, modified by CTH.

1.2. Voice Anatomy

Since this dissertation focuses on the voice source, this section is devoted mainly to the review of the anatomy of the larynx:

The larynx is an intrinsic component of the respiratory system. Its function is to act as a protective device for the lower respiratory tract. As a valve it (a) prevents air from escaping the lungs when needed; (b) prevents foreign substances from entering the larynx (and thus the lungs); (c) forcefully expels foreign substances which threaten to enter either the larynx or the trachea (Zemlin, 1998). Based in the order of sequential phylogenetic acquisition, the primary function of the larynx is sphincteric, protecting the lower airway from the intrusion of liquid and food. Its secondary function is respiration, and its third (and most recently acquired) function is phonation (Asher *et al.*, 1996).

Most commonly, the voice source of (pitched or non-pitched) sounds is constituted by the vibrating vocal folds. In some singing styles, other structures such as the ventricular folds, the ary-epiglottic folds or the arytenoid cartilages, can also constitute part of the vibrating system (Bailly *et al.*, 2010; Fuks *et al.*, 1998; McGlashan *et al.*, 2007; Sakakibara *et al.*, 2004).

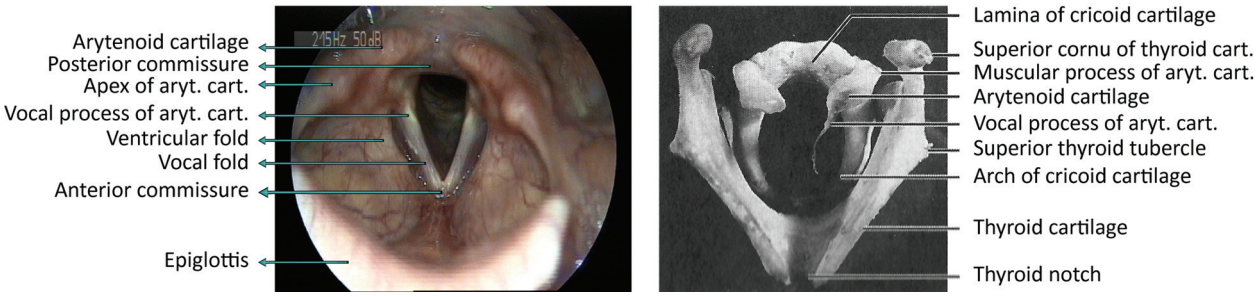


Figure 2 - Laryngeal anatomy. Left: Top view of larynx during aspiration (female, 18 years). As is customary in laryngology, the picture is oriented as if the person was lying on her stomach. Thus, the right vocal fold is displayed on the left side of the image, and vice versa; right: cartilages of the larynx, top view. Taken from Zemlin (1998), Fig. 3.12.

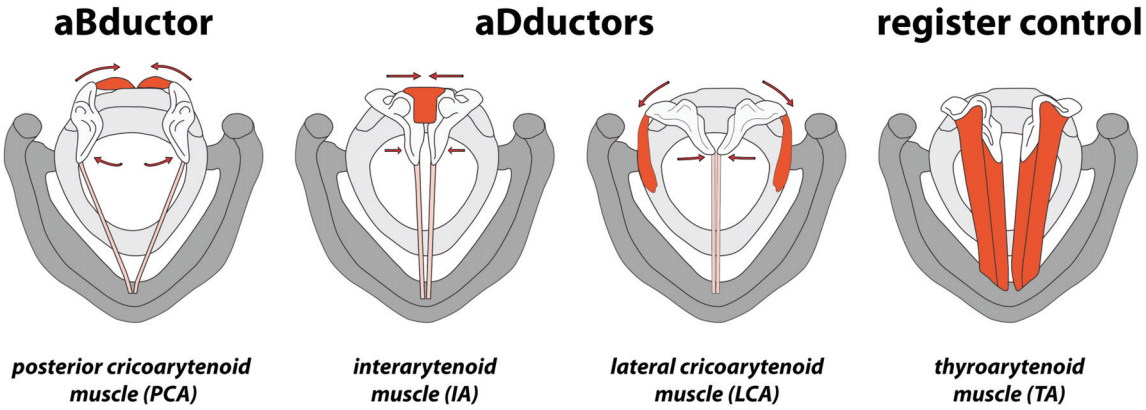


Figure 3 - Schematic illustration of four intrinsic laryngeal muscles: the PCA is a vocal fold aBductor; the IA and LCA are aDductors; TA thickens and medially bulges the vocal fold, thus (a) decreasing vocal fold length; (b) increasing intrinsic vocal fold tension; and (c) adducting the anterior portion of the vocal folds. The TA is mainly responsible for controlling the vocal register in singing – see chapter 1.6.3.1. Based on Zemlin (1998), Figures 3.52, 3.55, 3.53, and 3.47.

The vocal folds are paired structures originating from the thyroid cartilage at the anterior commissure, and inserting superiorly into the (paired) vocal processes of the arytenoid cartilages – see Figure 2. The glottis is the space between the vocal folds. The vocal folds (and thus the glottis) can be configured for either breathing (in which case the vocal folds are aBducted) or for phonation (the vocal folds are aDducted to a certain degree) via the rotation of the arytenoid cartilages. In particular, aBduction is facilitated by the posterior cricoarytenoid (PCA) muscle, and aDduction is maintained through the lateral cricoarytenoid (LCA) and the interarytenoid (IA) muscles – see Figure 3. An example of the reconfiguration of the glottis from breathing to phonation is shown in Figure 4.

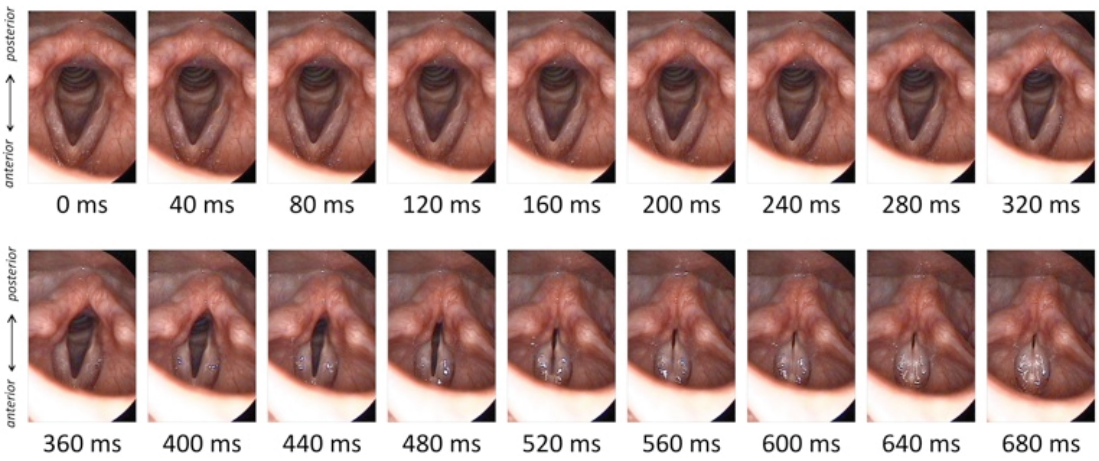


Figure 4: Videolaryngoscopic recording of an adductory gesture and phonation onset performed by an adolescent singer phonating at pitch Db3 (ca. 139 Hz). Phonation onset occurred at t = 520 ms.

Each vocal fold consists of muscle tissue, i.e. the thyroarytenoid (TA) muscle, and a vocal ligament, which is continuous with the conus elasticus (Zemlin, 1998) – see Figure 5. As a vibrating system, the vocal folds constitute a five-layer-system, consisting of the epithelium, three layers of the lamina propria (superficial, intermediate, and deep), and the muscle. These layers have different biomechanical properties, which is an important fact when discussing the registers of the singing voice (see further, paragraph 1.6.3.1).

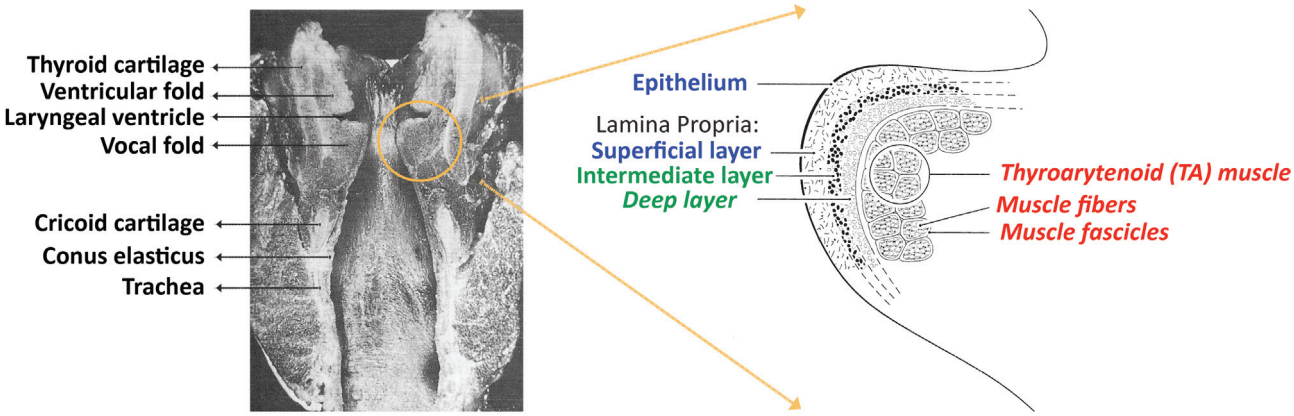


Figure 5 - Vocal fold structure. Left: frontal section of the larynx, taken from Zemlin (1998), Fig. 3.24; Right: schematic illustration of frontal section through the right vocal fold, showing tissue layers. Blue: mucosa; green: ligament; red: muscle. Vocal fold cover layers are labeled with regular characters; layers belonging to the vocal fold body are labeled with italics – taken from Titze (2000), Fig. 1.13. Note that Hirano (1981) proposed a slightly different classification system, where the "body" only consists of the vocalis (TA) muscle.

1.3. Capturing and visualizing voice

A wide range of methods and protocols has been developed in order to capture and document phonation (Baken and Orlikoff, 2000). Here, only the methods related to manuscripts **I – III** are briefly described.

1.3.1. Acoustic recording

Sound is an audible pressure wave (Rossing, 1990). In the case of speech and singing, it is (mostly) created by pressure waves radiating from the mouth and partly from the nose (recall Figure 1). These pressure waves can be recorded by a system consisting of a microphone and a microphone pre-amplifier, upon which they are converted into a time-varying (AC) voltage. Several issues need to be considered when choosing the proper equipment for acoustic recording, such as the microphone frequency response, frequency range, dynamic range, and directionality, as well as the load impedance of the dynamic range and the gain of the microphone pre-amplifier. Since a detailed discussion is beyond the scope of this dissertation, the reader is referred to the excellent tutorial by Švec & Granqvist (2010). This publication also discusses the effects of mouth-to-microphone distance on the recorded signal.

1.3.1.1. Recording equipment

In modern recording systems, acoustic data is stored digitally. In order to do so, the time-varying voltage coming from the pre-amplifier must be digitized by an analog-to-digital converter (ADC), i.e. a sound card. Such a device puts out a string of binary numbers at each period of the sample clock (a timer driven at a particular frequency, i.e. the sampling frequency) (Roads and Strawn, 1998). These binary numbers can then be stored in the processor's RAM (random access memory), or on a storage device like a hard disk or a CD. This process is called sampling, and it converts a continuous-time signal (i.e. an analog signal) into a discrete-time signal (i.e. a digital signal).

1.3.1.2. Sampling

According to the Shannon sampling theorem, a continuous-time signal x(t) with frequencies no higher than f_{max} can be reconstructed exactly from its samples $x[n] = x(nT_s)$, if the samples are (at least) taken at a rate $f_s = 1 / T_s$, which is greater than 2 f_{max} (McClellan *et al.*, 1998). With other words, an acoustic signal with the highest frequency component f_{max} can only be digitized correctly, if the sampling frequency f_s is at least twice f_{max}. If this principle is violated, the reconstructed digitized signal will contain frequencies that have not been present originally, and this phenomenon is called *aliasing*. (The principle of aliasing is employed for creating stroboscopic effects, such as in videolaryngostroboscopy – see chapter 1.3.2.2). Modern sound cards have a built-in safeguard against aliasing, since they remove all energy above half the sampling frequency (i.e. the so-called *Nyquist frequency*) with a low-pass filter (Roads and Strawn, 1998).

1.3.1.3. Frequency analysis

When describing the quality of a spoken or sung sound, one must distinguish between a *subjective* psychoacoustic level (i.e. the **voice timbre**) and an *objective* physical level (i.e. the **sound spectrum**). Voice timbre is formally defined as *that attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch as dissimilar* (ANSI, 1960). By this definition, its perception can vary from individual to individual (thus, timbre can not be measured objectively). The sound spectrum, on the other hand, is determined by the frequency and amplitude (and phase) of each component of a complex vibration (Rossing, 1990). In other words, it is *a physical property that can be characterized as a distribution of energy as a function of frequency* (Roads, 1996).

To obtain a sound spectrum, the acoustic signal can be decomposed into its individual sinusoidal components (each having a frequency, an amplitude, and a phase offset) with the Fourier transform. For an infinite continuous-time signal, the process is described in Equation 1. For a discrete-time signal of finite duration, this process is defined in Equation 2 (Smith, 2007).

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$$

Equation 1: Fourier transform of a continuous-time signal; t represents time in seconds, ω is the frequency, and X(ω) is a complex number representing amplitude and phase of each respective sinusoid.

$$X(\omega_k) = \sum_{n=0}^{N-1} x(t_n) e^{-j\omega_k t_n}$$

Equation 2: Discrete Fourier Transform (DFT) of a finite-duration discrete-time signal; $x(t_n)$ is the input signal value at t_n (sec); ω_k is the frequency; $X(\omega_k)$ is the complex-valued spectrum at frequency w_k; N is the number of time samples in the analyzed signal.

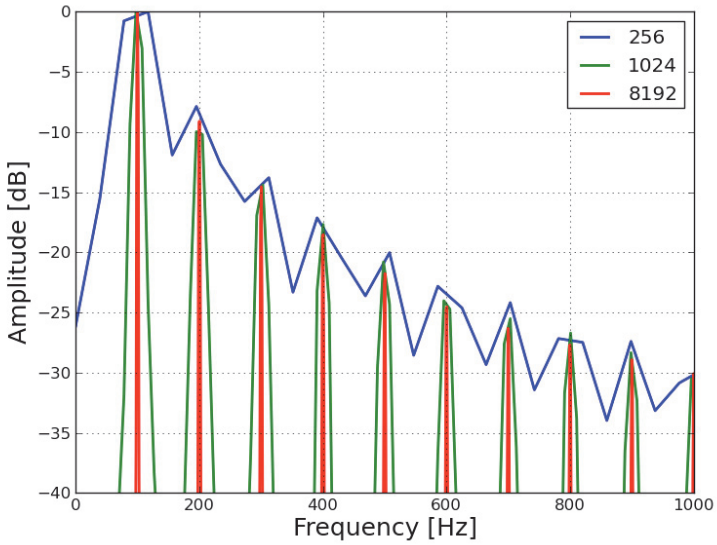


Figure 6 – Effect of FFT window size on the spectral representation of a simplistic model of the human sound source (harmonic series having a fundamental frequency of 100 Hz and a spectral flatness of -9 dB/octave[1], synthesized at a sampling frequency of 10000 Hz). The same signal was analyzed with three different FFT window sizes: blue = 256 samples (wide-band); green = 1024 samples; red = 8192 (narrow-band). Note that the wide-band representation creates the illusion of energy components between the individual harmonics (located at integer multiples of the fundamental frequency), which were in fact not present in the analyzed signal.

[1] See section 1.6.1 for a definition of spectral slope.

Based on Equation 2, the spectrum (consisting of amplitude and phase information for each frequency) of any digitized signal can be computed (for the purposes of voice analysis, the phase information is discarded most of the times). In a spectrum, frequency is plotted at the x-axis and amplitude on the y-axis. Note that time is discarded during the calculation of the spectrum. A spectrum provides information on energy components within a certain portion of the signal. The longer this analyzed portion of the signal, the finer the frequency resolution gets (i.e. the more individual frequencies can be detected within this portion of the signal), but at the cost of decreasing temporal resolution (i.e. it can not be said any more when, within the analyzed portion of the signal, a specific energy component occurred). In more mathematical terms, the spacing of the frequencies in the spectrum is inversely proportional to the length and duration of the analyzed portion of the signal, i.e. the *window size*. This tradeoff has to be considered when choosing between a *wide-band* (smaller window size, hence better time resolution and worse frequency resolution) and a *narrow-band* (larger window size, hence better frequency resolution and worse temporal resolution) spectrum and spectrogram. For more information refer to (Howard, 1998) – see also Figure 6.

1.3.1.4. Spectrogram

A spectrogram is a time-varying representation of spectral density – see Figure 7 for an example. Here, many spectra, calculated at equally spaced time intervals, are visually combined into a new display. In order to do so, each spectrum is "rotated": In a spectrogram, frequency is displayed on the y-axis, and amplitude is displayed on the z-axis (as color intensity). The time offset, at which each individual spectrum has been calculated, is displayed on the x-axis. Note that these processing steps are analogous to the new EGG wavegram technique, which is described as part of this dissertation in chapter 3.2.

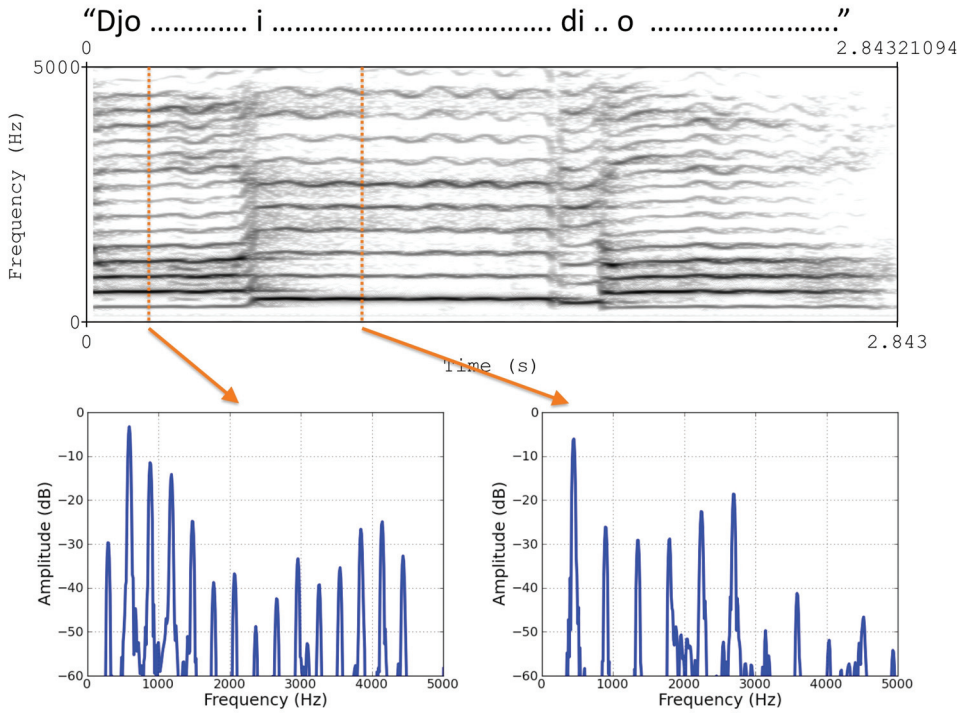


Figure 7 – Spectrogram of yodeling (microphone signal, first four notes of the "Mariazeller Jodler", female yodeler, 60 yrs). 1st row: narrow-band spectrogram; 2nd row: individual spectra extracted at t = 0.2 sec (left) and t = 1 sec (right).

1.3.2. Visual documentation of vocal fold vibration

1.3.2.1. Laryngoscopy

The larynx, including the vocal folds, is located in a confined space within the body. Hence, viewing the larynx during voice production is not trivial. The basic technique to observe the vocal folds is called laryngoscopy – see Figure 8 for a schematic illustration. Basically, three difficulties have to be overcome in order to perform a successful laryngoscopy: (a) the vocal folds do not lie in a straight line of sight, due to the curved nature of the vocal tract; (b) the vocal folds have to be properly illuminated; (c) the examined person may exhibit protective airway reflexes (Mallampati, 1996), thus obstructing the view of the vocal folds. The first documented (unsuccessful) attempts to visualize the vocal folds in a living human date back to 1807. Only about half a century later, researchers succeeded in acquiring visual data on the vocal folds with a quality good enough for scientific documentation (for historical overview see e.g. Moore, 1991).

Modern laryngoscopy is performed either with a flexible endoscope that is inserted through the nose, or with a rigid endoscope via the oral cavity. Whereas the flexible endoscopy has the advantage of permitting normal articulation in speech and song, it generally has a lesser image quality (Verikas *et al.*, 2009). The opposite is true for rigid endoscopy.

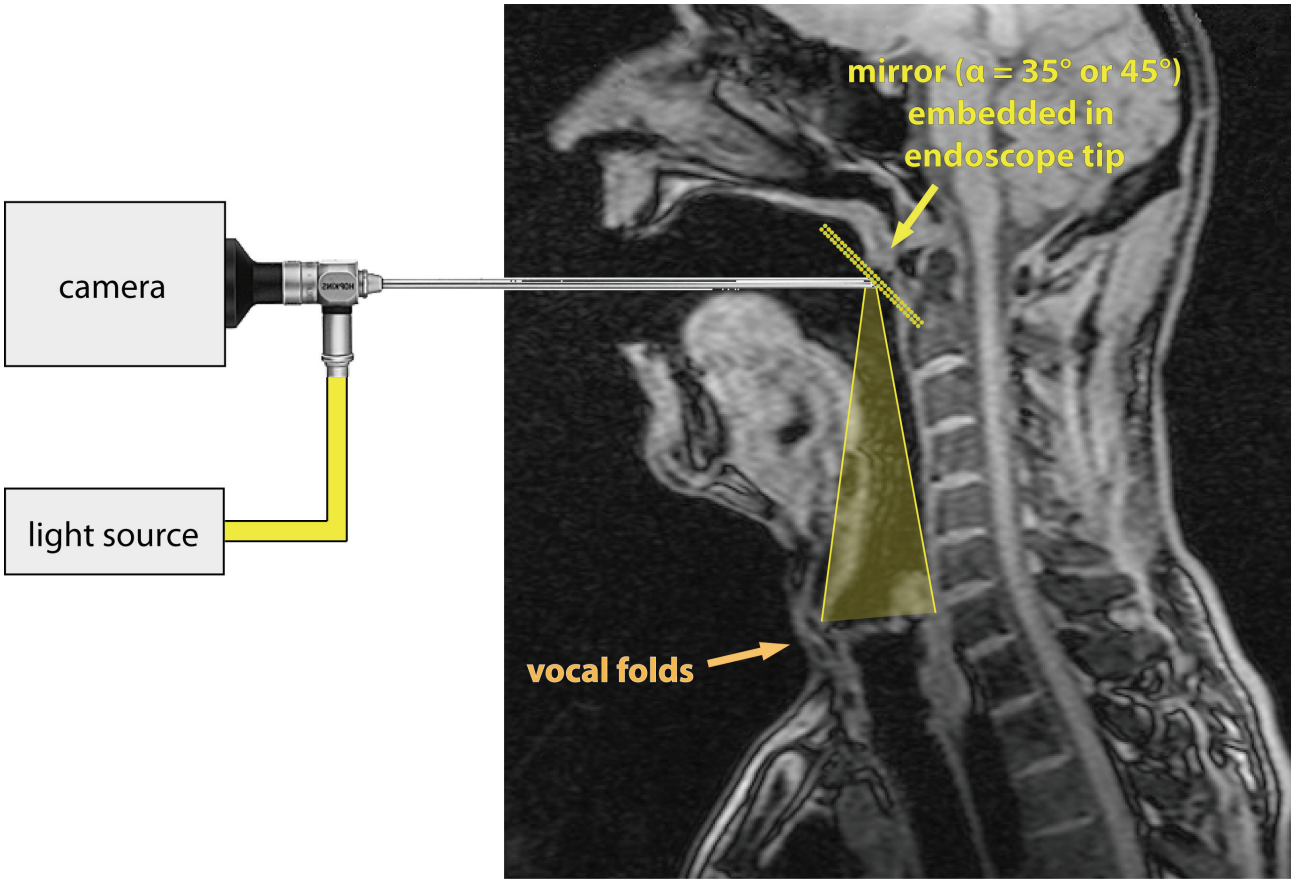


Figure 8 – Schematic illustration of the principle of laryngoscopy: A rigid endoscope is inserted into the oral cavity. Illumination is provided by an external light source. A mirror at the tip of the endoscope is used to deflect both the illuminating light source and the recorded image at the back of the pharynx. Different configurations of this method have been realized, utilizing (a) a flexible endoscope inserted through the nose; (b) a rigid endoscope with a camera sensor embedded at the tip of the endoscope.

1.3.2.2. Videostrobolaryngoscopy

The fundamental frequencies of human phonation are reported to lie in the range of 73 – 659 Hz for the speaking voice (Friedrich and Dejonckere, 2005), with extremes at ca. 40 and 2100 Hz (Baken and Orlikoff, 2000), and sometimes even higher (Neubauer *et al.*, 2004). Vocal fold vibrations at these frequencies are too fast (a) to be observed with the naked eye; and (b) to be recorded with consumer grade video cameras, which can only capture 25 – 30 images per second.

As a possible solution to this dilemma, the technique of stroboscopy can be employed, resulting in the method of videostrobolaryngoscopy when applied to the observation of the vibrating vocal folds in the larynx. In this method, the fundamental frequency of (nearly) periodic phonation is estimated in real-time, and a light source is triggered at a frequency of usually 1 or 2 Hz below the fundamental frequency of vocal fold vibration – see Figure 9. This effectively creates the illusion of vocal folds vibrating at 1 or 2 Hz, i.e. the frequency difference between the measured fundamental frequency and the frequency of the strobe flashes. One serious limitation of this method is that it only can be applied to (nearly) periodic vocal fold vibration. The technique will fail to capture voicing onsets or irregular vocal fold vibration.

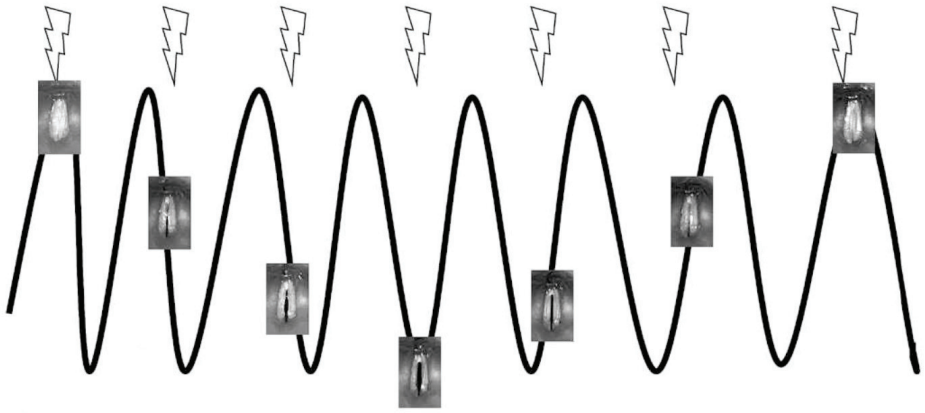


Figure 9 - Illustration of the principle of videostroboscopy: Strobe flashes are triggered at a frequency that is related, but not identical, to the fundamental frequency of (nearly-periodic) vocal fold vibration. Usually, the frequency difference Δf is 1 or 2 Hz, resulting in the illusion of slow-motion vocal fold vibration with an apparent fundamental frequency of Δf Hz. Image taken from Friedrich and Dejonckere (2005), Fig. 3a.

1.3.2.3. Videokymography (VKG)

For adequate documentation of the vibratory characteristics of each glottal cycle, multiple images must be captured per cycle. This implies that the rate of image acquisition must be well above the fundamental frequency of phonation. This is accomplished in high-speed video (HSV) recording. State of the art technology allows for image acquisition rates of 4000 – 10000 frames per second (Mehta and Hillman, 2008), with high-end devices going up to 20000 fps (Deliyski and Hillman, 2010).

Videokymography (VKG) is a low-cost alternative to HSV imaging (Švec and Schutte, 1996). A modified video camera allows for repeated scanning of one horizontal line (transverse to the glottis) at a rate of 7812.5 Hz. The successive line images (line height 1 pixel) are concatenated to form a videokymogram, displaying time on one axis (either going from top to bottom or from left to right, dependent on the orientation of the videokymogram). The time-

varying line images taken at the scanning line position are displayed on the other axis – see Figure 10 for details.

Videokymography allows to observe the following features of vocal fold vibration in a high temporal resolution: vertical phase differences of vocal fold vibration, mucosal waves, shapes of lateral and medial peaks, vibratory a-symmetries, and different types of irregularities and cycle aberrations (Švec *et al.*, 2007). One disadvantage of the technique is that it cannot capture anterior-posterior phase differences (so-called "zipper-like" opening and closure – Childers *et al.*, 1986; Hess and Ludwigs, 2000), since data is only captured at the position of the single scan line.

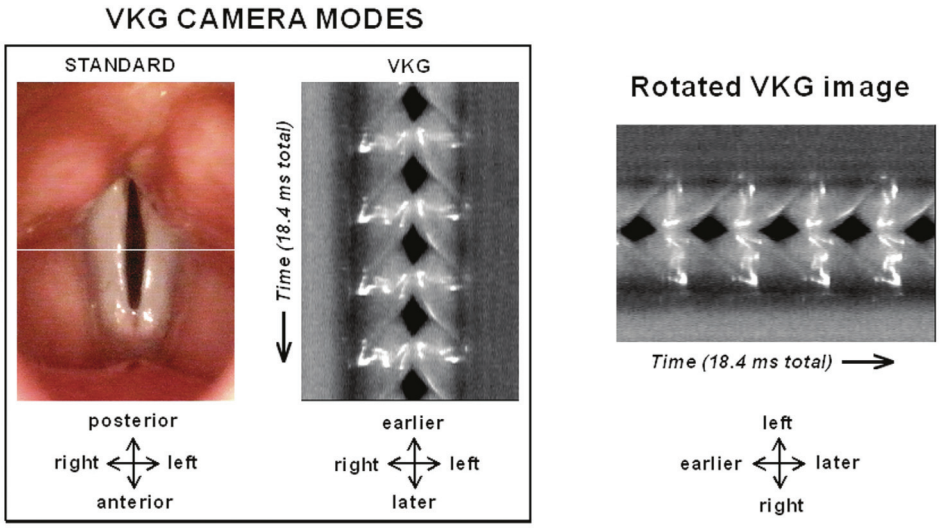


Figure 10 – Videokymography: two modes of the videokymographic camera. The measuring position for the VKG image is marked by the line in the standard image; the videokymogram displays the vibratory pattern of the middle part of the vocal folds. On the very right, the VKG image is rotated to show the time along horizontal axis. The investigated subject was a female with normal vocal folds phonating with moderate effort at the frequency of about 250 Hz. Taken from Švec *et al.* (2009), Fig. 15.2.

1.3.3. Electroglottography (EGG) – a physiological correlate of vocal fold vibration

The largest disadvantage of direct visual observation of the vibrating vocal folds is its invasiveness. Laryngoscopy can influence a person's habitual way of phonation - this particularly applies to rigid endoscopy, where the articulation of only a single vowel ("endoscopic vowel /i/) is possible. Laryngoscopy causes discomfort, and some people might not tolerate the procedure altogether. In order to circumvent these shortcomings, several methods of capturing correlates of vocal fold motion have been developed (Fabre, 1957; Hertz *et al.*, 1970; Holmer *et al.*, 1973; Rothenberg, 1977; Sondhi, 1975; Sonesson, 1960). One of these methods, electroglottography (EGG), is being advanced in this dissertation.

Electroglottography was introduced by Fabre (1957) to monitor the vibration of the vocal folds in vivo. A low intensity, high-frequency current is passed between two electrodes that are placed externally on the neck, on each side of the thyroid cartilage at vocal fold level. The contacting and de-contacting of the vocal folds causes variations in the electrical impedance across the larynx, resulting in a variation of the current between the two electrodes (Baken, 1992; Fourcin and Abberton, 1971) – see Figure 11. This time-varying current has been found to be related to changes in vocal fold contact area (Scherer *et al.*, 1988).

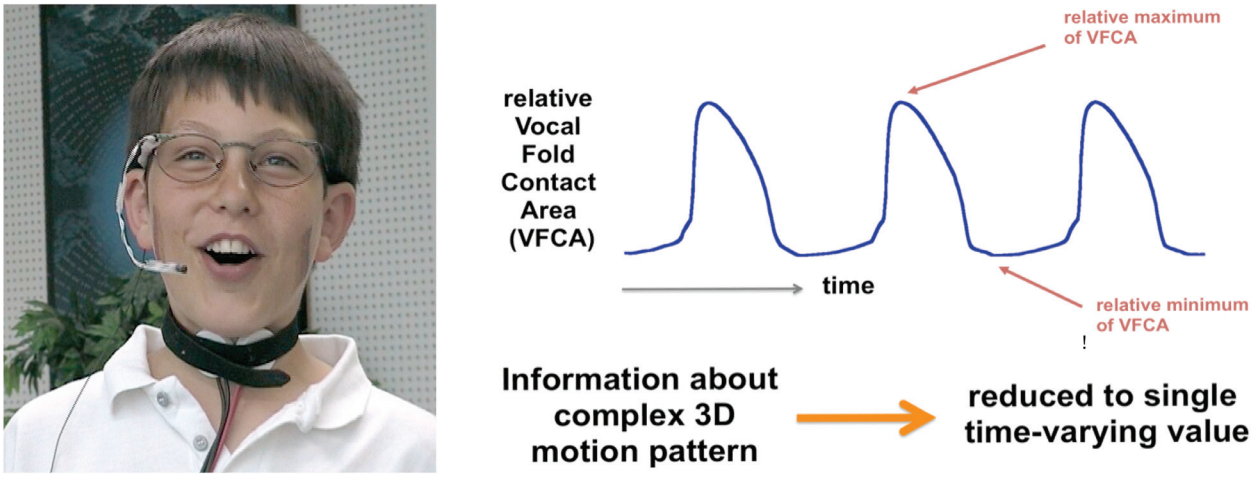


Figure 11 – Electroglottography (EGG). Left: paired EGG electrodes attached to the thyroid cartilage of a young singer; Right: typical electroglottographic signal (three glottal cycles displayed).

The electroglottographic signal is a time-varying one-dimensional representation of the complex three-dimensional motion of the vocal folds.

1.3.3.1. Landmarks in the EGG waveform

A word of caution is appropriate when considering the terminology used: The terms "closure", "closing", "opening", "closed phase" etc. relate to vocal fold vibration and to air flow. Electroglottography, on the other hand, measures relative vocal fold contact area, which is an altogether different quantity. The EGG signal conveys information about "contact", "contacting" and "de-contacting" events, and the "contact phase". The data derived from EGG might only be used as a rough physiological proxy for their counterparts in vocal fold vibration and airflow – see e.g. chapter 1.6.1.

It has been shown that landmarks in the EGG waveform are related to the movement and position of the vocal folds during phonation (Hess and Ludwigs, 2000; Rothenberg, 1979). The model suggested by Rothenberg is shown here, as represented by Baken and Orlikoff (2000) :

(a) Begin of the contacting phase. Increase of vocal fold contact commences, starting at the lower margins of the vocal folds.

(b) Vocal folds begin to contact fully with closure at their upper margins.

(c) Maximum vocal fold contact area reached for this glottal cycle, possibly coinciding with full closure of the vocal folds (particularly in chest register). Note however, that the EGG signal alone is not useful for assessing glottal closure, which might be incomplete, e.g. due to a posterior glottal chink (see chapter 1.6.2) or pathological changes of the vocal fold geometry.

(d) End of maximum contact and start of the opening phase from the lower margins of the vocal folds.

(e) Start of separation of the upper margins of the vocal folds.

(f) End of the opening phase and start of the open phase, indicating minimum vocal fold contact in the glottal cycle. It is safe to assume that the glottis is open during this phase of the glottal cycle, and glottal airflow is larger than zero.

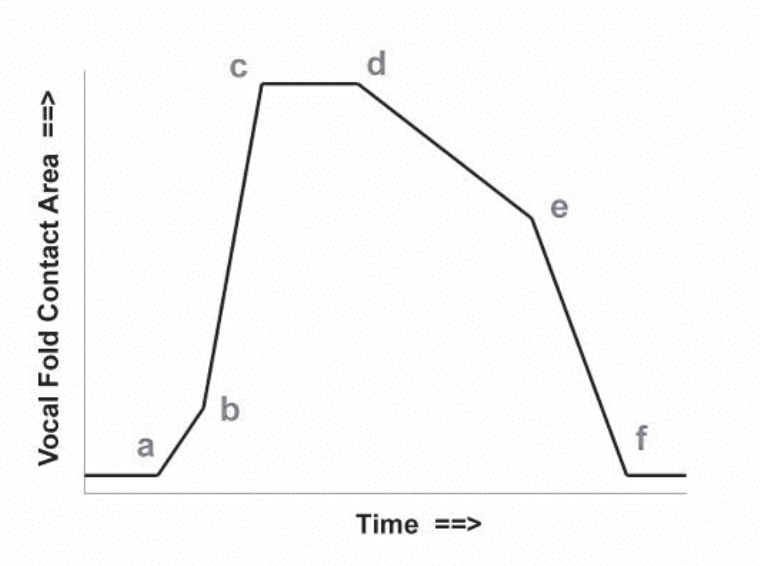


Figure 12 - A model representing the landmarks in the electroglottographic signal of one cycle of vocal fold vibration. (Baken and Orlikoff, 2000)

The physiologic relevance of the EGG signal has been examined theoretically by Titze (1989; 1990), who discussed the effects of (a) increased glottal adduction; (b) glottal convergence and vertical phasing; (c) medial vocal fold surface bulging, facilitating a so-called "knee" in the EGG waveform (see Figure 12e and, more importantly, the "chest register" case in Figure 22); and (d) increased vertical phasing in vocal fold vibration.

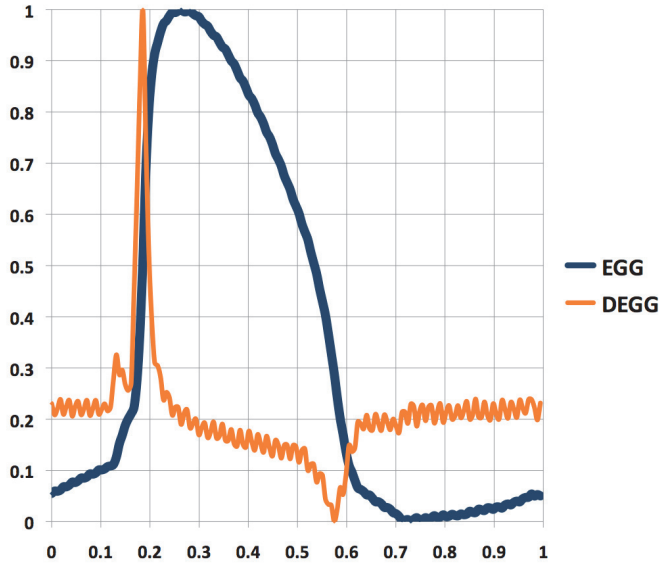


Figure 13 - EGG waveform (blue) representing one glottal cycle, and its first derivative, DEGG (orange). The DEGG maximum and minimum indicate vocal fold contacting and de-contacting events, respectively.

The first mathematical derivative of the EGG waveform (DEGG) reflects the rate of change of the EGG with time (Childers and Krishnamurthy, 1985; Teaney, 1980) – see Figure 13. Experimental research (Baer *et al.*, 1983; Childers *et al.*, 1983) has confirmed a close relation between peaks in the derivative of the EGG signal and the contacting and de-contacting events of the vocal folds. However, in some subjects, multiple peaks for both the contacting and the de-contacting phase can be seen. It has been suggested that the multiple peaks in the DEGG signal might be related to zipper-like opening and closing of the vocal folds (either anterior-to-posterior or posterior-to-anterior – Henrich *et al.*, 2004; Hess and Ludwigs, 2000). It is also conceivable that they result from artifacts, such as mucus strands (Colton and Conture, 1990), or abnormalities in vocal fold tissue structure distorting the regularity of the EGG waveform (Baken and Orlikoff, 2000; Kitzing, 1990). Such DEGG multiple peaks phenomena have, however, not yet been investigated in detail. One of the benefits of the newly developed method of EGG wavegrams (see chapter 3) is allowing a more thorough investigation of such phenomena.

1.3.3.2. The EGG contact quotient

Quantitative analysis of the EGG waveform has been done mostly by estimating the relative proportion of glottal closure within a glottal vibratory period (Rothenberg and Mahshie, 1988), known as the 'larynx closed quotient' (Howard, 1995) or 'contact quotient' (Orlikoff, 1991) (CQ_{EGG}). Rothenberg and Mahshie (1988) proposed a 'criterion-level method' to calculate the CQ_{EGG}: The peak-to-peak amplitude of each vibratory cycle is computed. The contacting event is defined as that point in time when the signal strength exceeds a certain 'criterion level' or 'threshold level', which is usually indicated as a percentage of the peak-to-peak amplitude. In past research, criterion levels between 20 % and 50 % have been used. The duration of the contacting phase is defined as the time interval between the de-contacting and the contacting events, and the CQ_{EGG} is computed as the duration of the contacting phase divided by the period duration – see Figure 14. Henrich *et al.* (2004), proposed a method called DECOM to calculate the CQ_{EGG} based only on the first derivative of the EGG signal (DEGG). A hybrid method for the calculation of the CQ_{EGG} has been proposed by Howard (Howard, 1995; Howard *et al.*, 1990). In this method, the contacting event is defined by the strongest (positive) peak in the DEGG signal during the contacting phase, whilst the de-contacting event is assumed to be the point in time when the EGG signal strength falls below a criterion level of ca. 42 % (three sevenths).

The contact quotient has been found useful in clinical as well as basic voice research (Schutte and Miller, 2001; Švec *et al.*, 2008). A sudden decrease of CQ_{EGG} has been reported, e.g., during the transition from chest/modal to falsetto register in singing (Henrich *et al.*, 2005; Miller *et al.*, 2002). Research has shown, however, that the CQ_{EGG} is dependent on the choice of algorithm used to determine the contacting and de-contacting events, and must therefore be used with caution (Herbst and Ternström, 2006; Higgins and Schulte, 2002; Kania *et al.*, 2004; Sapienza *et al.*, 1998) – see Figure 15. This shortcoming of CQ_{EGG} estimation has partially motivated the development of the EGG wavegrams, which are being introduced in this dissertation (see chapter 3).

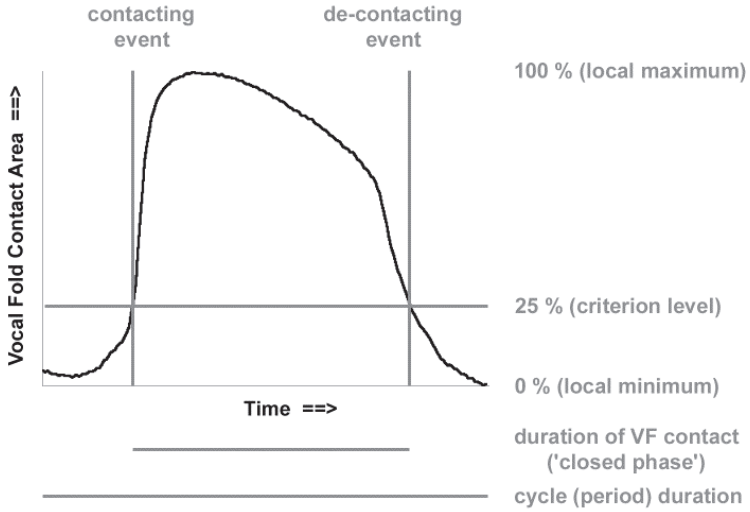


Figure 14 - Calculation of the EGG contact quotient by a criterion-level method. In this case, a threshold of 25% of the local signal peak-to-peak amplitude is used. Graph taken from Herbst and Ternström (2006), Fig. 3.

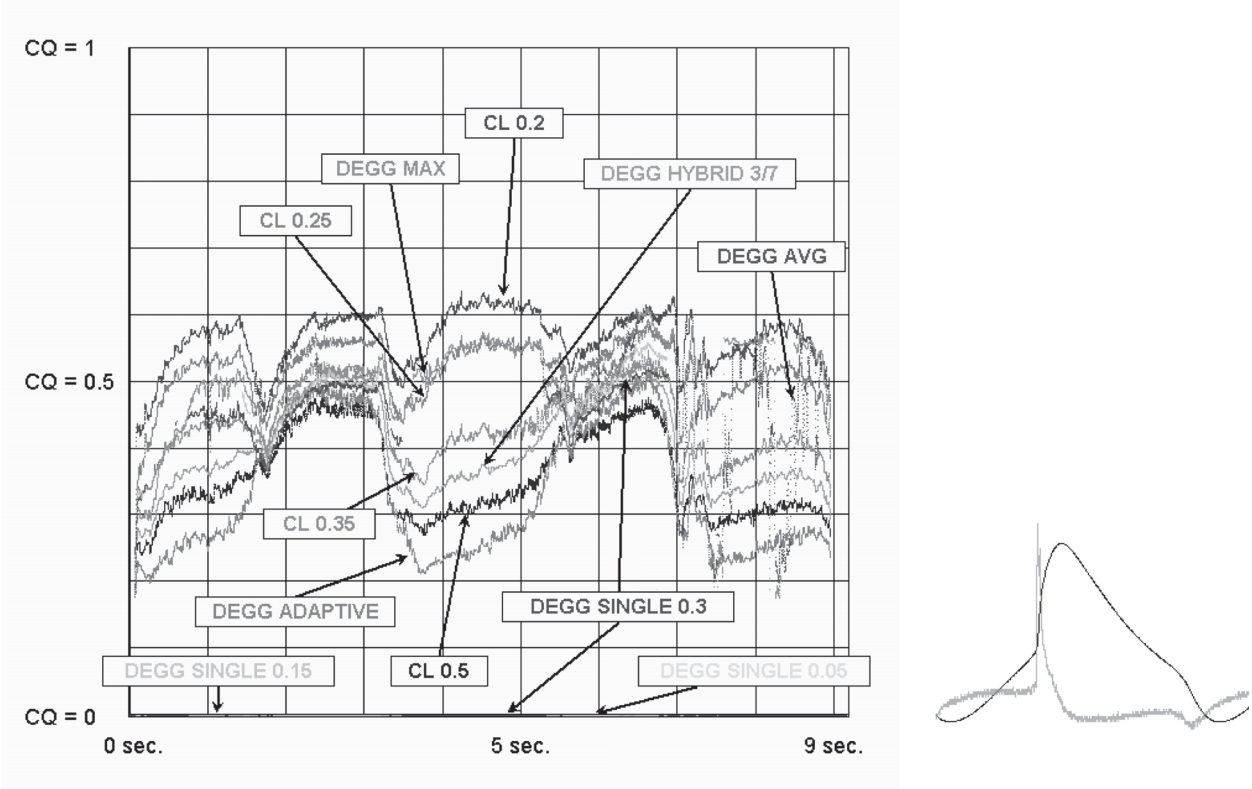


Figure 15 – Variation of CQ_{EGG} data, as calculated with different algorithms. Left: The CQ_{EGG} data for a series of octave glides from F#2 (185 Hz) to F#3, as calculated with different methods at a sampling rate of 10000 Hz. The region between 4 and 5 seconds in the audio signal, where the CQ_{EGG} values differ by as much as 0.3, represents the second repetition of the note F#2 after a downward glide from F#3; Right: EGG waveform (dark) and its 1st derivative, DEGG (light) of a glottal cycle extracted at t = 4.5 s. Data from Herbst and Ternström, (2006), Figures 5 and 6.

1.4. Source-filter theory

The source-filter theory of sound production has been established by Fant (1960). According to this theory, the radiated acoustic spectrum is created by a linear superposition of sound source spectrum (created by the time-varying glottal flow), vocal tract resonances (formants), and mouth radiation. In other words, "the sound source is most often generated by the vibrating vocal folds that convert the steady (DC) airflow from the lower respiratory system into a periodic train of flow pulses. These pulses, referred to as the glottal flow, are then acoustically filtered by the vocal tract resonances. This process effectively redistributes the amplitudes of the frequency components of the source signal, resulting in a vowel sound." (Story, 2002) – see Figure 16. This linear theory is useful for describing basic phenomena like vowel timbre of low-pitched sounds (i.e. in speaking). It allows the analytic separation of sound source and vocal tract, in order to study these systems separately. However, particularly at higher fundamental frequencies (when the frequency of the first harmonic exceeds the frequency of the first formant), the coupling of sound source and vocal tract might become non-linear (see chapter 1.6.1 for details).

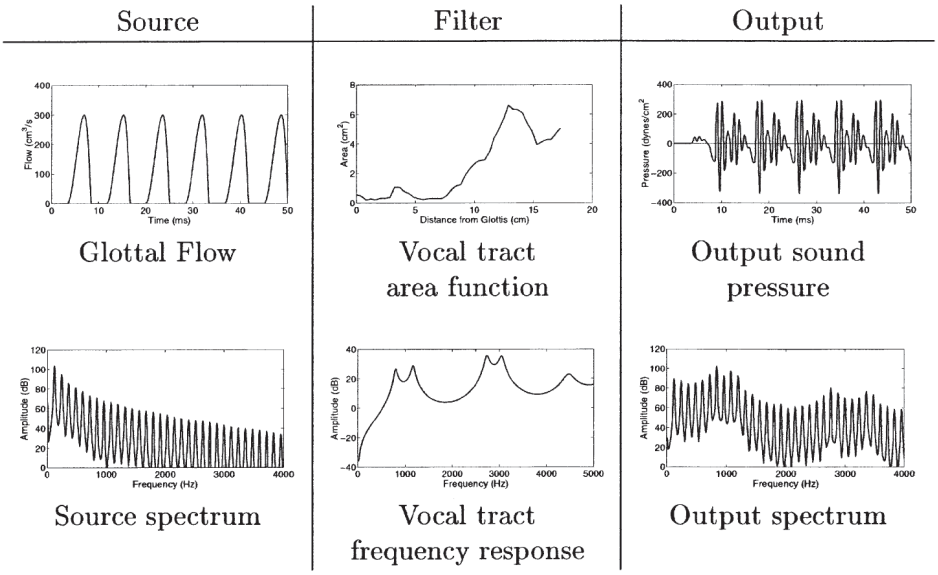


Figure 16 - Illustration of the source-filter representation of vowels. The upper row shows the source waveform, a vocal tract area function, and the output waveform. The second row shows the frequency domain representation of each of the quantities in the first row. Graph from Story (2002), Fig. 1.

1.5. Myoelastic-aerodynamic theory of voice production

The primary sound source in speech and singing is created by flow-induced self-sustaining oscillations of laryngeal tissue, driven by air coming from the lungs. In other words, the steady (DC) tracheal airflow is converted into a time-varying glottal flow. This phenomenon has been described in the myoelastic-aerodynamic theory of voice production (van den Berg, 1958). Here, vocal fold closure is facilitated by (a) a time-varying glottal shape: *convergent* (as seen in the frontal plane) in the opening phase, allowing maximum energy transfer from the air stream into the tissue; and *divergent* in the closing phase, creating a drop of intraglottal pressure that aids vocal fold closure; and (b) an inert supraglottal air column (i.e. a delayed vocal tract response caused by an inert acoustic load), also aiding in the closure of the vocal folds (Titze, 1988b).

This basic mechanism of voice production is fundamentally different from other sound production mechanisms seen in mammals, such as purring in cats. In purring, phonation is caused by a centrally driven periodic laryngeal modulation of respiratory flow (Sissom *et al.*, 1991): The intermittent activation of intrinsic laryngeal muscles (induced by a very regular, stereotyped pattern of muscular activity occurring ca. 20–30 times per second), causes glottal closure and the development of a trans-glottal pressure which generates sound when dissipated by glottal opening (Remmers and Gautier, 1972). Despite early claims (Husson, 1950), there is no evidence of such a sound production mechanism in humans (van den Berg and Tan, 1959).

In comparison with the rate of movement during possible vibratory frequencies of the vocal folds (ca. 40 to over 2000 Hz), the speed at which the intrinsic laryngeal muscles move is relatively small – according to Hunter *et al.* (2004), cyclic vocal fold posturing movements in humans can only be achieved up to ca. 10 times a second. Therefore, the *active* muscular adjustments of the vocal folds (i.e. adduction and choice of register, see chapters 1.6.2 and 1.6.3.1) are not to be confused with the much faster *passive* vocal fold oscillations, which are governed by aerodynamic and mechanical principles.

Nevertheless, as is shown in this dissertation, the configuration of the vocal folds via intrinsic laryngeal muscles has a basic influence on the vibratory properties of the vocal folds. Changes introduced into the vibration of the vocal folds are translated into changes of the acoustic spectrum via the time-varying glottal flow. In order to better understand this causal connection, it is necessary to consider glottal flow in more detail, as is done in the next section:

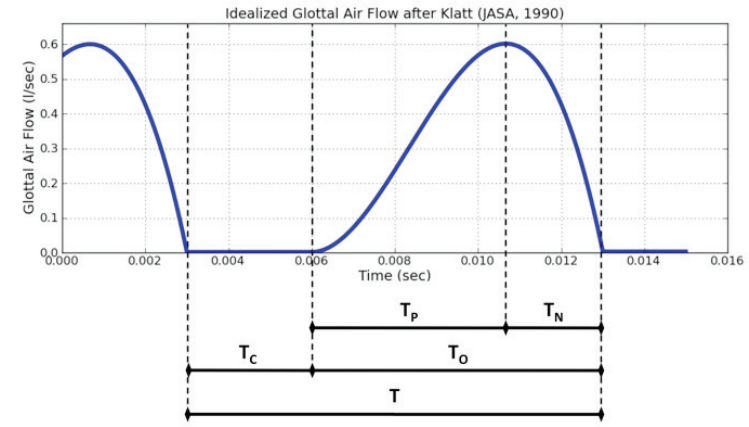
1.6. Modifying sound source timbre at the glottal level

1.6.1. Air flow and the voice source spectrum

The vibrating vocal folds open and close more or less periodically, allowing pulses of air to pass from the trachea into the pharynx. Interactions between air and tissue result in synchronized tissue velocity and glottal air pressure variations. The interruption of the flow at the instant of vocal fold closure sets up damped acoustic oscillations in the supraglottal as well as in the subglottal system of the trachea, bronchi, and lungs (Fant, 1979). The period T_0 (i.e. the duration of the glottal cycle) at which the glottal pulses repeat is directly related to the fundamental frequency (F_0) of phonation, whereas the detailed shape of these airflow pulses has an impact on the quality (and hence the timbre) of the created sound.

Based on the time-varying glottal airflow, each glottal cycle can be divided into (a) a (quasi) *closed phase* where the glottis (the space between the vibrating vocal folds) has reached its minimum during the glottal cycle. In this phase of the vibratory cycle, the airflow has either totally stopped due to full vocal fold contact, or air escapes at a steady rate if there is incomplete glottal closure; (b) an *open phase* where air is flowing - see Figure 17. The ratio between closed phase and the period is called the closed quotient (Q_C or CQ), and the ratio between the open phase and the period is the open quotient (Q_O or OQ). The open phase can conceptually be divided into a portion where the airflow increases (commonly called "*opening phase*"), and a portion where the airflow decreases ("*closing phase*"). In most cases of phonation, the closing phase is shorter than the opening phase, resulting in a skewing of the glottal waveform (Rothenberg, 1973). The ratio between the opening and the closing phases of the glottal pulse can be expressed as the *skewing quotient* (Q_S). The open quotient

and the skewing quotient are important quantities, since they define the shape of the glottal waveform (Doval *et al.*, 2006; Titze, 2000) – see below.



T_C = time of absence of airflow ("closed phase")
T_P = time of increasing airflow ("opening phase")
T_N = time of decreasing airflow ("closing phase")
T_O = duration of flow ("open phase")
T = period of oscillation

Q_O = open quotient = T_O / T
Q_C = closed quotient = T_C / T = 1 - Q_O
Q_S = skewing quotient = T_P / T_N

Figure 17 - Idealized glottal flow waveform, synthesized with the model described by Klatt & Klatt (1990).[2]

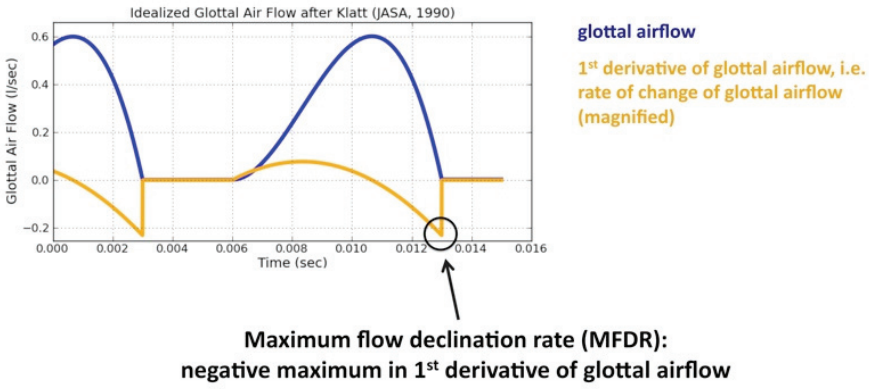


Figure 18 - Relation of glottal airflow to its 1st derivative, i.e. its rate of change. The negative maximum of the 1st derivative is the maximum flow declination rate (MFDR). The MFDR occurs at the point in time where, caused by abrupt cessation of glottal airflow during glottal closure, most of the acoustic energy is being created during a glottal cycle.

[2] The model was calculated as $U_k(t) = at^2 - bt^3$, where $b = \frac{27AV}{4T_0^2 O_q^3}$, $a = \frac{27AV}{4T_0 O_q^2}$, AV is the amplitude of voicing, T_0 is the period, and O_q is the open quotient.

The rapid termination of glottal airflow at the instant of vocal fold closure in each glottal cycle creates strong high-frequency components that excite the higher order vocal tract resonances (Miller, 1959; Rothenberg, 1973; Sundberg, 1981b). This is particularly the case if the closing phase is abrupt, i.e. the first derivative of the glottal flow (the rate of change of the airflow) has a strong negative peak at the moment of vocal fold closure, resulting in a large *maximum flow declination rate* (MFDR) (Holmberg *et al.*, 1988) – see Figure 18. The MFDR (and thus the timbre of the sound) is heavily influenced by both the open quotient and the skewing quotient (Fant, 1979).

The distribution of energy in the source spectrum, i.e. the amplitude of the individual harmonics, decays uniformly in the frequency domain with increasing frequency (Rothenberg, 1981b). This energy distribution is termed *spectral slope* and is measured in negative decibels (dB) per octave. The spectral slope depends on the sharpness of the cessation of airflow, becoming less negative (i.e. flatter) the more abrupt it is. Typical spectral slope values would range from -6 dB/octave for *brassy* phonation (with more high-frequency components in the spectrum) to -18 dB/octave for *flutey* (without noise components) or *breathy* phonation (with noise components in the glottal source, created by turbulent airflow in the insufficiently adducted glottis – Fant, 1979; Titze, 2000), see Figure 19. The model shown in Figure 19 suggests that there is a direct relationship between the spectral slope and the abruptness of glottal closure (which is partly influenced by the duration of the closed phase, and hence the closed quotient).

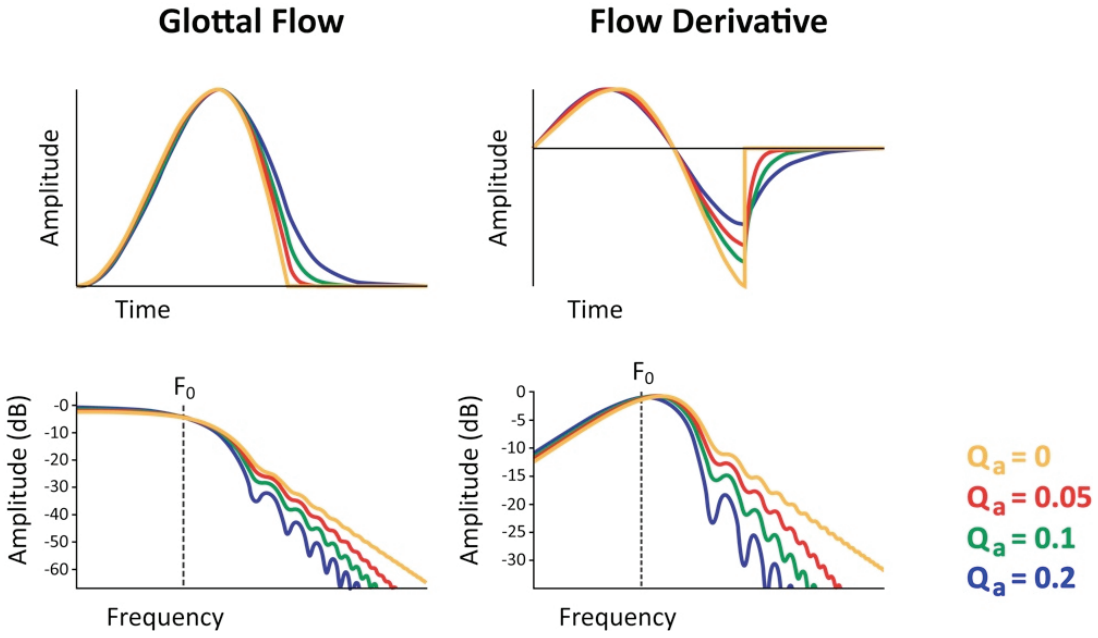


Figure 19 - Effects of asymmetry of airflow waveform on created spectrum. Top row: waveforms of glottal flow (left) and its derivative (right) for a generic glottal flow model. The varied parameter is the return phase quotient(Q_a), which is defined as the ratio between the return phase time constant[3] and the duration between the glottal closure instant (GCI) and the end of the period. Bottom row: energy spectra of the glottal flow waveform (left) and of its derivative (right). Note that a smaller Q_a (and hence a smaller open quotient) results in a flatter spectral slope. Data from Doval *et al.* (2006), Fig. 13. In a nutshell, the data suggests that the spectrum contains stronger high-frequency components (i.e. it has a flatter spectral slope) as the closed quotient increases.

[3] In the generic glottal flow model, the return phase quotient is formally defined as: $Q_a = T_a / [CqT_0]$, where T_a is the time constant of the exponential decay in the return phase, C_q is the closed quotient, and T_0 is the period (Doval *et al.*, 2006).

A vocalist has two main options for influencing the shape of glottal flow waveform and thus the timbre of the sound source: (a) by introducing changes into the quality of vocal fold vibration and glottal air flow through adjustments of intrinsic laryngeal musculature, which will be discussed in the reminder of this chapter; (b) by increasing the vocal tract inertance by narrowing the epilarynx tube or by moving a (subglottal or supraglottal) formant (i.e. an energy maximum in the vocal tract transfer function) just above the fundamental frequency (Rothenberg, 1981a; b; Titze, 2004a; b).

1.6.2. Vocal fold adduction

The relevance of glottal adduction in singing has already been observed more than 160 years ago by Manuel Garcia:

'When one very vigorously pinches the arytenoids together, the glottis is represented only by a narrow or elliptical slit, through which the air driven out by the lungs must escape. Here each molecule of air is subjected to the laws of vibration, and the voice takes on a very pronounced brilliance. If, on the contrary, the arytenoids are separated, the glottis assumes the shape of an isosceles triangle, the little side of which is formed between the arytenoids. One can then produce only extremely dull notes, and, in spite of the weakness of the resulting sounds, the air escapes in such abundance that the lungs are exhausted in a few moments.' (Garcia, 1847b)

In this statement, Garcia describes how the voice timbre can be influenced via controlling a physiologic parameter - the adduction of the posterior glottis. In addition, Garcia hints at the inverse relation between glottal adduction and airflow rates, and the fact that the maximum phonation duration is dependent on the degree of glottal adduction.

The degree of closure in the posterior part of the glottis can be increased by contraction of the LCA and IA muscles (Baken and Isshiki, 1977; van den Berg *et al.*, 1960; Broad, 1968; Fried *et al.*, 2009; Hunter *et al.*, 2004; Letson and Tatchell, 1997; Sataloff, 1997; Zemlin, 1998) – recall chapter 1.2.[4] An incomplete adduction of the cartilaginous portion of the vocal folds can be described laryngoscopically as a "posterior glottal chink" (PGC - Södersten *et al.*, 1995). Variation of the cartilaginous adduction is an important modulator in both speech and singing, i.e. in the production of breathy, normal, and pressed or creaky voice (Ladefoged, 2001; Zemlin, 1998). Whilst in breathy voice the arytenoid cartilages are set apart, in pressed voice they are usually squeezed together.

The degree of vocal fold adduction has a major influence on glottal airflow. Figure 20 shows typical time-varying airflow rates for four types of phonation: "breathy", "normal", "flow" and "pressed" (Sundberg, 1981a). Acoustically, the most important feature is the steepness of the waveform at the closing phase of the glottal cycle, indicating how fast the airflow is terminated. This parameter is closely related to the MFDR (see chapter 1.6.1): a glottal flow waveform with a steeper closing phase will result in an acoustic output with a larger degree of high-frequency energy (and hence a flat spectral slope), as is e.g. the case in "flow" phonation. In "breathy" phonation, on the other hand, the closing phase of the glottal flow waveform has a more gradual slope, thus suggesting that the created sound will be less brilliant.

[4] During phonation, a supplementary role in arytenoid adduction/abduction is played also by the thyroarytenoid (TA) and cricothyroid (CT) muscles, for details see Zemlin (1998).

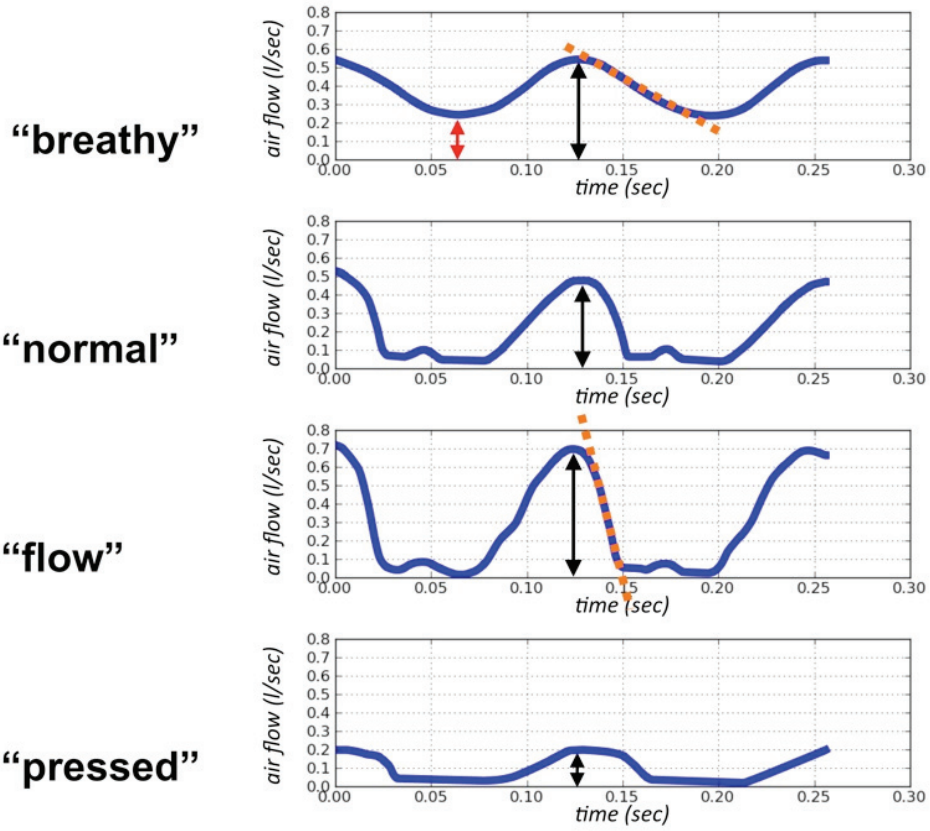


Figure 20 - Airflow rates for one cycle of phonation in four different qualities: "breathy", "normal", "flow" and "pressed" (data was obtained by direct measurement and subsequent inverse filtering of the air flow waveform). The pulse amplitude and the closed phase duration both change for different adduction settings. Note that in the "breathy" case the airflow never goes below 0.2 l/sec (red arrow), indicating incomplete glottal closure. The rate of change in the closing phase (i.e. the steepness, illustrated here by dashed orange lines for the two extremes, i.e. "breathy" and "flow" phonation) varies considerably depending on phonation type, having an effect on the steepness of the spectral slope of the acoustic output. Data taken from Sundberg (1981a).

The overall airflow, and hence the amount of air used per unit of time, can be determined by averaging the airflow graphs in Figure 20 over time. It appears that "flow" phonation, whilst being the most efficient type from a spectral point of view, uses in overall less air than "breathy" phonation. Consequently, maximum phrase duration is increased. This is in direct agreement with Garcia's quote from the beginning of this chapter[5].

[5] Sundberg's model consists of three ideal phonation types ("breathy", "flow" and "pressed") and focuses on the glottal wave shape and the MFDR. Garcia's quote, on the other hand, is based on direct laryngeal observation, and suggests a dichotomy of aBducted vs. aDducted phonation. These two models can only be related to each other with caution: It is unclear whether Garcia's "vigorous pinching of the arytenoids" is equivalent to "flow" or "pressed" phonation, or some mixture thereof.

1.6.3. Registers in singing

Vocal registers are usually described by different authors using one or more of the following approaches:

- the singer's kinaesthetic awareness of vibrations during sound production - e.g. *voce di petto* (chest voice) and *voce di testa* (head voice), suggested in the late 16th century by L. Zacconi (Stark, 1999)
- perceptual qualities, such as timbre – see e.g. Titze (1988a)
- vocal tract (or pitch) adjustments, i.e. changes in the relation of the frequencies of individual harmonics to the frequencies of individual vocal tract formants – see e.g. Miller (2000) for an example of "full head" register in male classical singing.
- different vocal fold vibratory regimes – see e.g. Garcia's classical description (1847b).
- individual teaching systems based on pedagogic experience, e.g. Chapman (2006) or Sadolin (2008).

The terminology used for describing vocal registers is vast: Mörner *et al.* (1963) report more than 100 terms used in the literature. An overview of register theories proposed in the last 150 years was compiled by Henrich (2006). Reid (2005) quotes the register definitions of nine contemporary (groups of) authors.

For the purpose of this manuscript, we define a vocal register as *a set or range of serial sounds that are similar in perception and produced by similar vocal fold vibratory patterns* (Sakakibara, 2003). Based on this definition, we recognize four registers (see Figure 21):

Vocal fry[6] (also: pulse register, strohbass, M0) is characterized by a fundamental frequency below 70 Hz (i.e. pitch C#2), where glottal pulses are perceived as localized bursts of acoustic energy followed by gaps of silence, rather than as a continuous sound: Each glottal excitation pulse is dampened by the vocal tract and the trachea, thus the acoustic output goes to zero before the next pulse occurs (Titze, 1988a).

The **chest** (also: modal, M1) register is closely related to speech, and is usually found in the low to mid portion of the singer's voice range, with ranges for females from about F3 (ca. 175 Hz) to F4 (ca. 349 Hz) with a "belting extension" to D5 (ca. 587 Hz); and for males from E2 (ca. 82 Hz) to C5 (ca. 523 Hz) (Miller, 2000)[7]. The chest register is usually characterized by a rich, sometimes *brassy* timbre.

Falsetto (also: head, loft, M2) register is located in the mid to upper portion of the voice range in humans, with ranges from about C4 (ca. 262 Hz) to F6 (ca. 1397 Hz) in females and A3 (220 Hz) to G5 (ca. 784 Hz) in males (Miller, 2000). Some females employ the falsetto register for speech purposes. The falsetto register is characterized by a *flutey* sound, hinting at a steeper spectral slope as compared to chest register.

Whistle register (also: flute register, flageolet, M3) is a phonation type that occurs, depending on authors, above pitches B5 (ca. 988 Hz) (Švec *et al.*, 2008), between G5 (ca. 788 Hz) and D6 (ca. 1175 Hz)(Miller and Schutte, 1993), or F6 (ca. 1397 Hz) (Titze, 2000).

[6] Whereas vocal fry is not used in classical singing, it is regularly used in contemporary commercial music (particularly as an effect at note onset).

[7] The pitch ranges indicated for different registers vary slightly from author to author.

Phonations in whistle register are reported to be created by tightly stretched and slightly abducted vibrating vocal folds (Švec *et al.*, 2008) or by jet-induced vortices (Berry *et al.*, 1996; Herzel and Reuter, 1997; Tsai *et al.*, 2006).

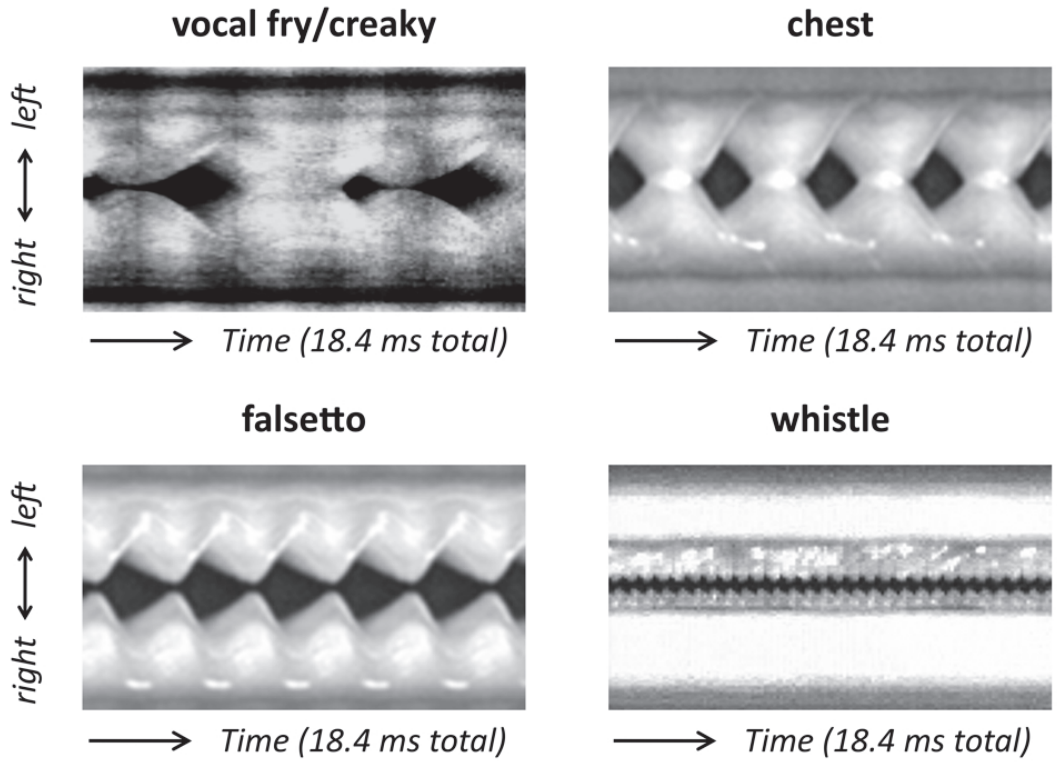


Figure 21 - Videokymograms of four registers in singing. The measurement line was placed in the middle of the vocal folds, transversal to the glottal axis. Note the qualitative differences in the vibratory characteristics: a) creaky voice/vocal fry: double opening, long closed phase, vertical phase differences (i.e. visible phase difference between the vibration of the upper and lower margins of the vocal folds); b) chest/modal voice: relatively long closed phase, vertical phase differences; pronounced mucosal wave; c) falsetto voice: no closed phase[8], slightly reduced vertical phase differences[9]; d) whistle register: no vocal fold contact, very limited vocal fold vibratory amplitude. Total time displayed: 18.4 ms in each case. Data from (Herbst *et al.*, 2011; Švec *et al.*, 2001; Švec *et al.*, 2008)

Vocal fry and whistle register can only be produced at the extreme ends of the human fundamental frequency range and are not typically employed during singing performance, since their fundamental frequencies lie outside the typical pitch ranges for females and males[10]. Phonations in chest and falsetto register are central to most singing styles, and consequently much more is known about them. We will therefore limit our discussion to these two registers.

[8] Note that both chest and falsetto register can be produced either with glottal closure, or without, depending on vocal fold adduction (see further in the text).

[9] Note that in some people phonating in falsetto register, the vertical phase difference can be either greatly reduced or even missing altogether.

[10] For a counter-example, see e.g. Neubauer *et al.* (2004).

1.6.3.1. Chest vs. falsetto register

Physiologically, the main difference between chest and falsetto register lies in the action of the thyroarytenoid (vocalis) muscle (TA):

In **chest register**, the TA is active, thus thickening, shortening and medially bulging the body of the vocal fold, whilst slackening the vocal fold cover (Hirano, 1974; Titze, 2000). In that way, a vertical phase difference in vocal fold vibration is created: The inferior portion of the vocal fold *leads* the vibration, and the superior portion *follows* with a phase delay of about 60° – 90° (Baer, 1975)[11]. This phenomenon asserts itself in the form of a mucosal wave (Hirano *et al.*, 1981), i.e. an airflow-driven travelling wave within the surface of the vocal fold tissue, moving from the inferior to the superior portion of the vocal folds once every glottal cycle. It is facilitated by the layered structure of the vocal folds, since these layers have different biomechanical properties: the *body*, consisting of the thyroarytenoid muscle and the deep layer of the lamina propria; and the *cover*, consisting of the epithelium and the superficial and intermediate layer of the lamina propria (Titze, 2000) – recall Figure 5. Both body and cover can move as separate units in vocal fold vibration (Hirano, 1974). Mucosal wave speed is inversely related to vertical phase delay. A strong, relatively slow mucosal wave, as seen in chest register, will both help to stabilize vocal fold vibration (a) by aiding the energy transfer from the air stream into the tissue during the opening phase of vocal fold vibration (Titze, 1988b); and (b) by prolonging the closed phase of vocal fold vibration. From an acoustic point of view, a longer closed phase (and thus a larger closed quotient) will enhance the output of high-frequency energy (Flanagan, 1958): In the open phase of vocal fold vibration, a considerable portion of the high-frequency energy generated by vocal fold vibration (and carried by the vocal tract's formants) is dampened. A longer closed phase will thus aid in creating a "brighter" voice timbre (Rothenberg, 1973).

In **falsetto register** both vocal fold body and cover are stretched (Hirano, 1974). This is accomplished by the action of the cricothyroid (CT) muscle (situated between cricoid cartilage and the thyroid cartilage, thus narrowing the cricothyroid space and elongating the vocal folds by tilting the thyroid cartilage forward (Zemlin, 1998). Here, the TA is relaxed or only slightly active. Due to a higher longitudinal tension in the vocal folds, their vibration is smaller in amplitude but higher in frequency. A mucosal wave is less likely to be seen, because the superficial vocal fold layer is under increased tension (Hirano, 1974), both reducing the vibratory amplitude and increasing the mucosal wave speed. This effectively shortens the closed phase, and no or little vertical phase difference is found in the tissue vibration. As a consequence, the vocal fold oscillation is more likely to be influenced by the vocal tract settings (Titze, 2008). The relatively longer open phase will cause less abrupt changes of air flow, which acoustically translates into a steeper spectral slope, thus limiting the high-frequency energy output (i.e. a more "flutey sound") – see Figure 22 for a comparison of chest and falsetto register phonation.

[11] In more recent research, a greater spread of vertical phase difference values has been reported (Jiang 2008, Jiang 1998)

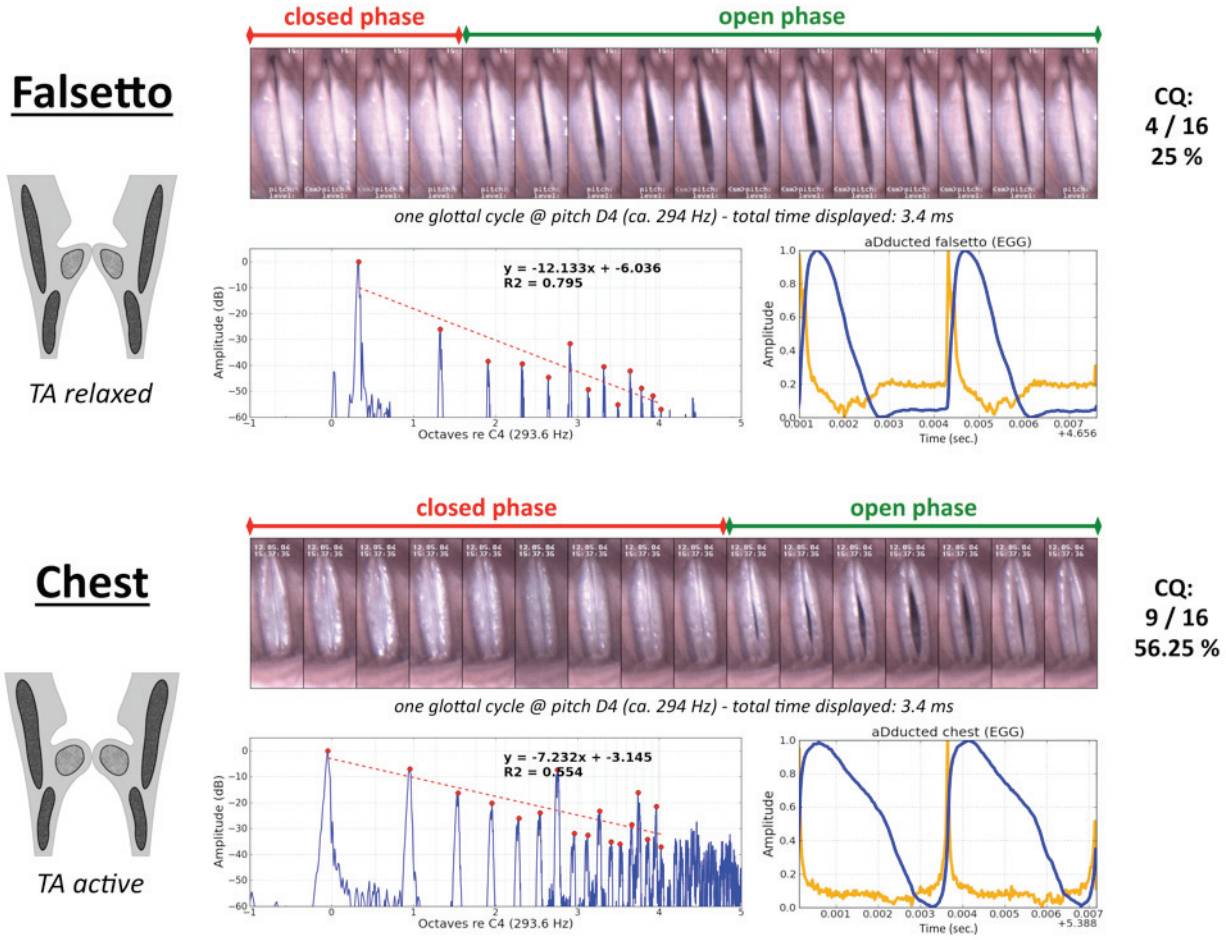


Figure 22 - Comparison of phonation in chest and falsetto register (male singer phonating with 'endoscopic' vowel /i/ at pitch D4 – ca. 294 Hz). Data displayed for or each of the two registers: schematic illustration of TA muscle activity in coronal section of larynx; videostrobolaryngoscopic images displaying one complete glottal cycle; acoustic spectrum with estimated spectral slope; two cycles of the EGG (blue) and DEGG (gold) waveform. The chest register phonation has, in comparison with the falsetto register: more bulged vocal folds; a longer closed phase; a flatter spectral slope (i.e. stronger high-frequency partials); a "knee" in the electroglottographic (EGG) waveform – see chapter 1.3.3.1 for details.

1.6.3.2. Blending the registers

Some singing styles (such as yodeling) are characterized by abrupt (voluntary) register transitions, resulting in abrupt timbral changes. In other styles (like classical singing and some sub-types of musical theater singing), these abrupt timbral and register changes are to be avoided, since they would violate the style's aesthetic principles. This is reflected in Lilli Lehmann's quite extreme postulate: "In singing pedagogy, registers should neither exist nor should they be created. As long as the term 'register' is used, registers will not vanish."[12] (Felsner, 2008), p. 63. Nawka and Wirth (2008) suggest that in trained classical singers, registers (and their boundaries and transitions) should not be perceivable at all.[13]

[12] "In der Ausbildung der Stimme sollen Register weder existieren noch geschaffen werden. Solange der Begriff verwendet wird, werden auch Register nicht verschwinden". – English translation by author CTH

[13] "Beim ausgebildeten Sänger dürfen die Register nicht als solche erkennbar sein. Bereits beim tiefsten Ton des Stimmumfanges müssen die Klangcharakteristika des obersten Tonbereichs mitschwingen.", p. 99

There is an obvious discrepancy between the physical and physiological reality (voice registers do indeed exist) and the aesthetical boundary conditions imposed by certain singing styles (registers and their transitions should not be audible). It is apparent that accomplished singers must have a means to navigate between chest and falsetto register without introducing abrupt timbral changes when changing either fundamental frequency or vocal intensity.

Many authors have argued for the existence of a so called “mid register” or “voix mixte”, which Seidner and Wendler (2004) [14] call into question as a “pedagogical fiction” and a mere “principle of operation”. The physiological groundwork has been laid by van den Berg (1963), who called the “mid register” not an "independent" register, but a "mixture" of chest and falsetto register: In chest register the longitudinal tension lies mainly in the muscular portion of the vocal fold, and in falsetto register the tension is mainly borne by the vocal ligament. The mid register is a “mixture of longitudinal tension in vocal muscles and vocal ligaments”, resulting in a “medium number of partials”. A skilled singer can navigate between chest and falsetto without producing breaks, by fine-tuning the action of the cricothyroid muscle[15] vs. thyroarytenoid muscle, and with the aid of minute adjustments of arytenoid adduction and subglottal pressure (van den Berg, 1963). With other words, the middle (or mixed) register can be defined as “a mixture of qualities from various voice registers, cultivated in order to allow consistent quality throughout the frequency range” (Sataloff, 1998, p. 370).

In contrast to these observations, some authors maintain that voix mixte sounds are always clearly produced in a given laryngeal “mechanism” (Castellengo *et al.*, 2004), i.e. either M1 (chest) for males and M2 (falsetto) for females. Voix mixte is supposedly not the result of an intermediate laryngeal process, unlike what the acoustic characteristics would suggest (Roubeau *et al.*, 2009). The necessary adjustments for unifying chest and falsetto register would, according to this theory, have to be made solely in the vocal tract.

In the light of the observations made by Miller and Schutte (2005), it would seem that both strategies (adjustments at the laryngeal level and resonance adjustments by changing vocal tract geometry) can be employed to successfully navigate the transition between chest and falsetto register. In fact, the voice as a vibrating system is the result of a complex interaction between three sub-systems, each having their own physical properties: the vocal folds (Švec *et al.*, 2000); and the supra- and subglottal vocal tract (Titze, 2004b; 2008). Each of these components has a general tendency to vibrate in its own set of frequencies (so-called eigenmodes). Since these systems are coupled, they are in constant “negotiation” with each other. Depending on the properties of the individual sub-systems, vocal fold vibration is either stable or tends to exhibit abrupt changes when the singer introduces changes in pitch or vocal intensity – see e.g. Švec *et al.,* (1999).

[14] “Das sogenannte *Mittelregister* wird in Publikationen derart unterschiedlich definiert, dass eine Stellungnahme schwerfällt. Muß dieses Register nicht als eine überwiegend gesangspädagogische Fiktion, als ein Bekenntnis oder Arbeitsprinzip aufgefasst werden, um die stets vorhandenen Register mit Aufmerksamkeit und Konsequenz an- oder auszugleichen?“, p. 103

[15] elevating the cricoid arch and depressing the thyroid lamina, and thus actively lengthening the vocal folds and increasing their tension.

2. Motivation for the work of the author

In speech and singing the sound is in most cases created by vibrating tissue in the larynx, i.e. in a concealed space. The direct visual inspection of the larynx during sound production is only possible since the mid-nineteenth century (Behnke, 1886; Garcia, 1847a), and key principles of voice production have only been understood in the last 50 years (e.g. Fant, 1960; Hirano, 1974; Ishizaka and Flanagan, 1972; Rothenberg, 1981a; Sundberg, 1974; Titze, 1983; van den Berg, 1958). The human voice (and in particular the singing voice) remains, therefore, an under-researched area.

Of the many possible projects in researching the (singing) voice, two have been chosen for this dissertation. Here, the motivations for these choices are elucidated.

2.1. Sound source documentation and analysis

Many methods to document speech and singing have been advanced, each having their particular advantages and disadvantages. When looking for a non-invasive, low-cost method for documenting vocal fold vibration, electroglottography is an obvious choice (see chapter 1.3.3). The technique is more that half a century old, and it is well established in research (Baken and Orlikoff, 2000). As a more recent development, its usefulness in voice pedagogy is being recognized (Herbst *et al.*, 2010; Howard *et al.*, 2004; Nair, 1999; Rossiter *et al.*, 1996).

Unfortunately, analysis of the EGG signal is not straightforward. Apart from possible problems with hardware and data acquisition protocol (Baken, 1992; Colton and Conture, 1990), there is a major issue that needs to be addressed: In order to study longer sequences of speech and singing, hundreds of glottal cycles have to be examined, especially when investigating phonations with changing physiologic parameters. Particularly in longer sequences, there is a danger of missing important but volatile information, or to loose the "big picture" altogether.

The need of extracting meaningful information from large amounts of EGG data is commonly addressed by calculating derived parameters, such as the EGG contact quotient (CQ_{EGG}) – see chapter 1.3.3.2. A previous investigation of this author (Herbst and Ternström, 2006) has shown, however, that the calculation of CQ_{EGG} data is dependent on human intervention (by the choice of algorithm in the first place, and secondly by setting analysis parameters). Also, the calculation of the CQ_{EGG} is conceptually unsuitable for EGG signals where there is no vocal fold contact, such as breathy (low-intensity) falsetto phonations (Herbst and Ternström, 2006). Additionally, there is a danger of overlooking important physiological information at the signal processing stage, when derived data is calculated, and the original EGG signal is disregarded for analytical purposes.

It is therefore desirable to look for an alternative method of analyzing the EGG signal, which conveys more information than just one time-varying analysis parameter. Such a method should provide the means to:

(a) visualize many consecutive glottal cycles in one graph

(b) retain the original appearance of the EGG waveform, or a close resemblance thereof.

2.2. Sound source adjustments in singing

The voice timbre is the central quality in singing: On a global or long-term level, singers (and their teachers) are concerned with "building the instrument", i.e. by establishing motor control and behavioural patterns which allow voice production within the limits of "acceptable" or "beautiful" singing in their chosen style. (This process is nicely reflected by the German term *Stimmbildung.)* On the other hand, singers must also be able to vary voice timbre on an ad-hoc level, allowing for artistic expression. Here, muscular fine-control and agility are paramount qualities of a "good" voice. On this matter, vocal pedagogue and singer Richard Miller (1986) suggests: "*Vocal timbre that results from a well-formed, well-coordinated instrument, without maladjustment of any of its physical parts or function, stands for the best chance of qualifying under the artistic criteria for tonal beauty, as found in Western culture.*"

The most obvious method of choice for voice timbre modification is via the sound modifiers (Howard and Murphy, 2007), i.e. articulation by introducing changes into the vocal tract shape (Ladefoged, 2001). Fine-tuning the vowel color (in relation to pitch) is a well-established concept for varying the vocal timbre and optimizing sound output (Coffin, 1980; Echternach *et al.*, 2010; Echternach *et al.*, 2011a; Henrich *et al.*, 2007; Joliveau *et al.*, 2004; Miller, 2008; Miller and Schutte, 2005; Story, 2004; Sundberg, 1972; 1974; 1977; Sundberg and Skoog, 1997; Wendler, 2008). Already in the mid-nineteenth century, however, M. Garcia (1847a; 1894) suggested that the sound quality can be influenced independently both via adaptations of the vocal tract and via modifications at the laryngeal level. Garcia's far-sighted observations have had an impact on many pedagogic schools of thought, such as Barbara Doscher's *Functional Unity of the Singing Voice* (1994). Doscher suggests that the voice timbre can either be influenced by adjusting the resonance tube, i.e. the vocal tract through change of the antagonistic vertical positions of soft palate and larynx (*bright* vs. *dark* quality), or in the larynx by the firmness of glottal closure (*ringing* vs. *veiled* quality).

At the laryngeal (i.e. sound source) level, the vocal timbre can be modified by two basic physiological adjustments:

(a) through changes in (posterior) vocal adduction, along the dimension of "breathy", "flow" and "pressed" (recall chapter 1.6.2); and

(b) through the choice of vocal register (i.e. chest vs. falsetto – recall chapter 1.6.3).

Whereas each of these physiological parameters is well researched and understood individually, little is known about their relation to each other. This is particularly true in the realm of voice pedagogy, where the proper terminology is not well established and sometimes not well comprehended.[16] It is therefore required to fill this gap in our understanding of the singing voice, by attempting to relate these two concepts (i.e registers and adduction) to each other.

[16] Out of politeness and respect for colleagues who mean well, I do not supply any reference here.

3. Original work by the author

3.1. Aims of this thesis

The aim of this thesis is twofold:

1. to develop a new method for visualization and analysis of the electroglottographic signal, i.e. a physiological correlate of vocal fold motion; This method should allow to display longer EGG signals in one graph, whilst retaining the appearance of each individual glottal cycle's waveform, as found in the original EGG signal.
2. to document physiological means of sound source control via adjustments of the vocal folds, by investigating the relation between voice registers and (posterior) glottal adduction.

Accordingly, this section of this thesis is divided into two parts: one concerning ***electroglottographic wavegrams*** (see manuscript I); and one dealing with ***laryngeal control of sound source properties*** (see manuscripts **II** and **III**).

3.2. Paper I - Electroglottographic Wavegrams

3.2.1. Methods

An algorithm has been developed by the author of this dissertation, to show in one single image (called an "electroglottographic wavegram") how the EGG waveform changes over time, from cycle to cycle. Figure 23 displays the process of wavegram generation in a nutshell:

(a) Consecutive glottal cycles of an EGG signal are extracted (Figure 23a).

(b) The locally normalized data values are converted into grayscale color information, and are plotted as a strip representing one glottal cycle each (Figure 23b).

(c) All strips are rotated 90 degrees counter-clockwise, and glottal cycle duration is normalized by scaling the individual glottal cycle plots to the same height (Figure 23c).

(d) The resulting graphs are combined to form the final display, the **EGG wavegram** (Figure 23d).

In the wavegram, time is displayed on the x-axis from left to right; the progress of consecutive glottal cycles is displayed on the y-axis from top to bottom; and the locally normalized vocal contact area is encoded in the z-axis as a color intensity value. The wavegram algorithm is described in detail in manuscript **I**. It was originally implemented with a C++ signal processing library written by this author. The code has later been ported to Python, and is in this form included as a supplement of this dissertation.

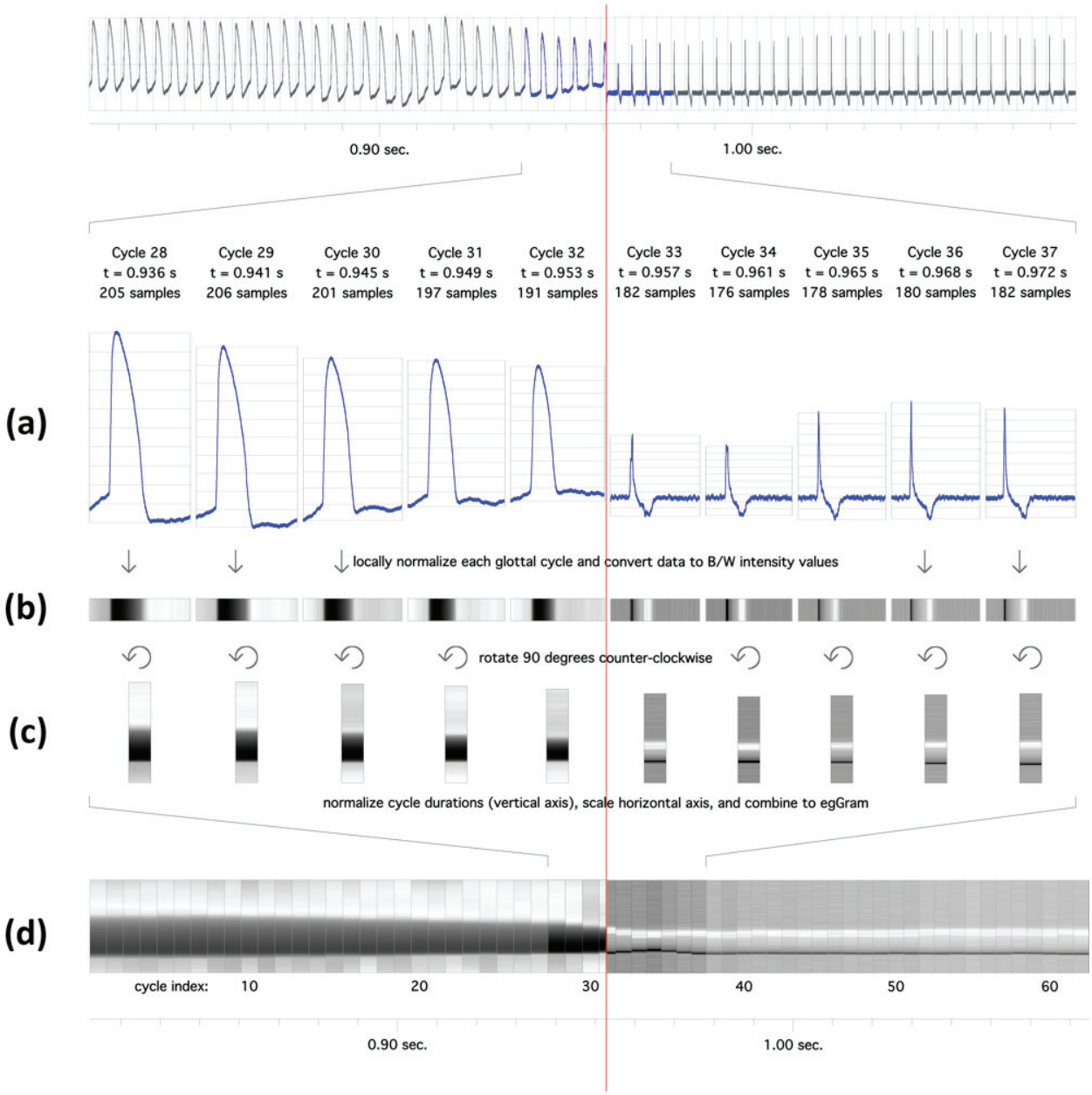


Figure 23: Basic processing steps to create an electroglottographic wavegram (see text). The two halves of the image illustrate the creation of wavegrams based on the EGG (left) and the DEGG (right) signal, respectively.

As an alternative display option, the first derivative of the EGG signal (DEGG) can be used as an input (rather than the EGG signal itself) to the algorithm of generating the wavegram – see the right half of Figure 23. DEGG-based wavegrams provide a clearer view of the moments of most rapid change in the vocal fold contact area and allow to track the potential double peaks in both the contacting and de-contacting phases of the DEGG-signal (Henrich *et al.*, 2004; Hess and Ludwigs, 2000).

The resulting wavegram can be "read" like book or a newspaper that has been rotated counter-clockwise by 90 degrees: In the rotated book (as in the wavegram), overall time goes from left to right, and each line in the rotated book is now displayed going from the bottom to the top.

3.2.2. Results

The typical landmarks found in EGG wavegrams are shown in Figure 24. In particular, they consist of:

1. start of glottal cycle, arbitrarily chosen in order to display the entire contact phase of the glottal cycle;
2. initiation of vocal fold closure;
3. maximum rate of increase of vocal fold contact;
4. maximum rate of decrease of vocal fold contact;
5. end of glottal cycle.

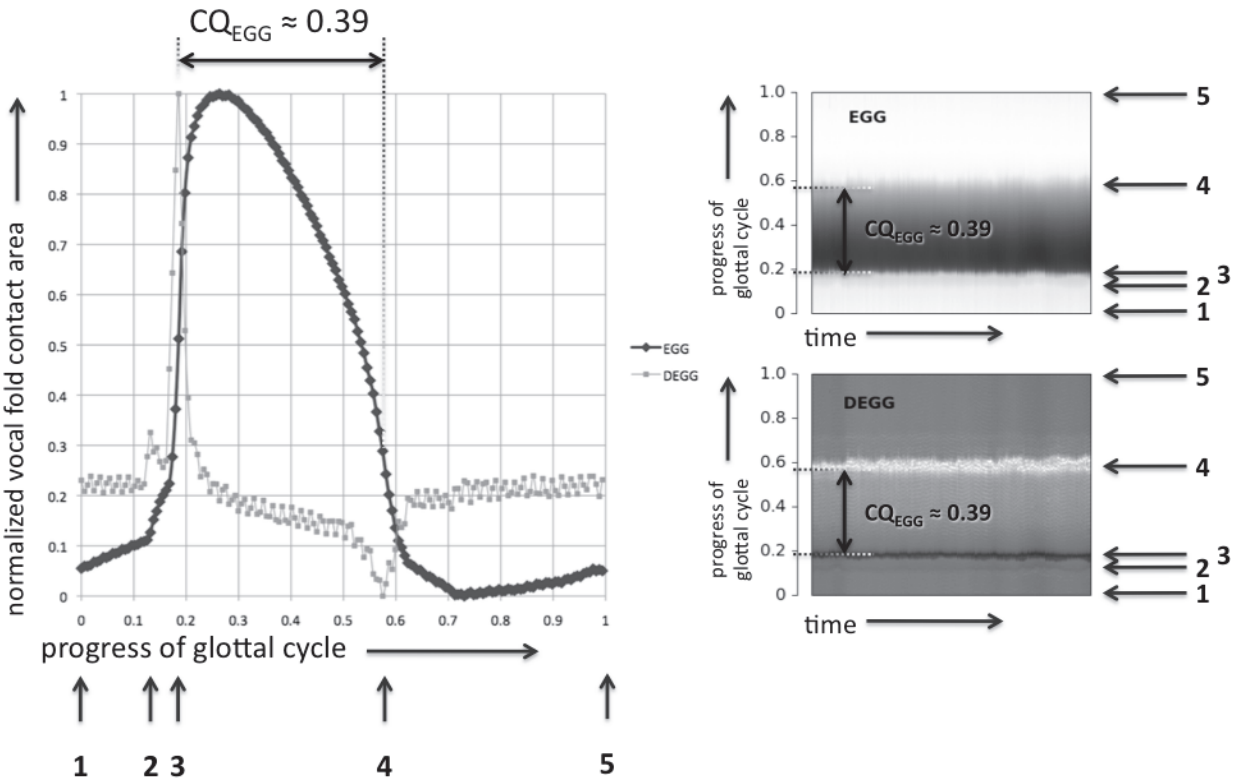


Figure 24: Typical landmarks in the EGG wavegram, both EGG-based [top right] and DEGG-based [bottom right], as related to the EGG signal of a single glottal cycle (dark) and its first derivative (light color).

These features agree well with the landmarks found in EGG signals (Childers and Krishnamurthy, 1985; Rothenberg). The temporal locations of the strongest positive and negative peaks of the DEGG signal can be used to estimate the EGG contact quotient (Henrich *et al.*, 2004). When applying the DEGG criterion to the glottal cycle displayed in Figure 24, a contact quotient of ca. 0.39 is calculated. The DEGG wavegram shows this clearly, since the positive and negative peak of the DEGG signal are displayed as distinct lines of darker and lighter color, respectively.

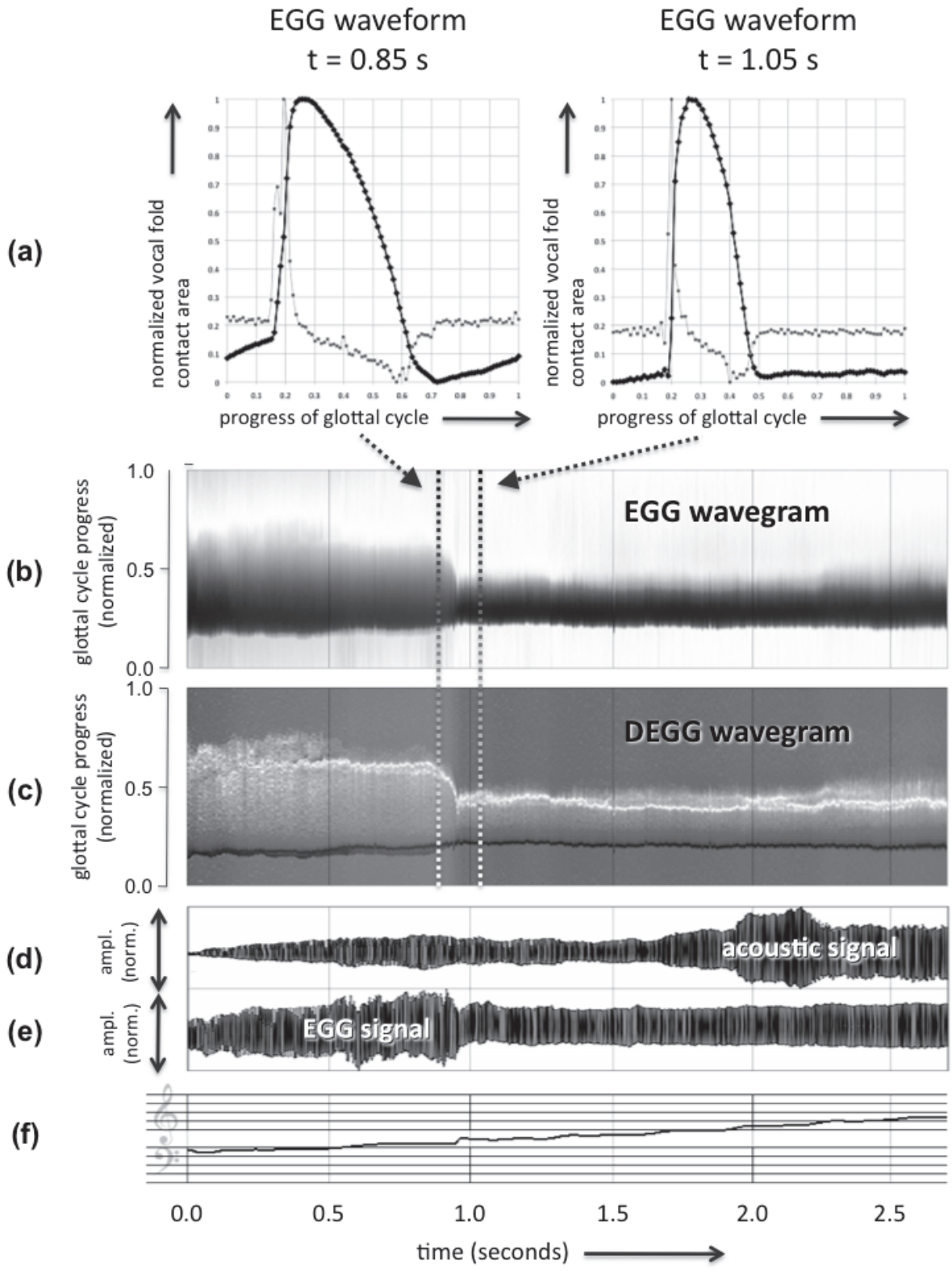


Figure 25: Changes of the EGG and audio signals in case of female phonation with increasing fundamental frequency. a) EGG (black) and DEGG (grey) waveforms representing glottal cycles extracted at t = 0.85 s and t = 1.05 s, respectively; b) EGG wavegram; c) DEGG wavegram; d) and e) amplitude plot of audio and EGG signal; f) fundamental frequency displayed in musical notation (ca. 208 Hz – 415 Hz).

In Figure 25, phonation with increasing fundamental frequency (ca. 208 – 415 Hz) is illustrated in a wavegram (female classical singer, 2 years of academic vocal training). Analysis data revealed a sudden change of EGG contact quotient around t ≈ 0.95 s, suggesting an abrupt (involuntary) break from chest to falsetto register. This coincided with (i) a sudden change of vocal fold contact phase (EGG waveforms and wavegrams, Figure 25a, b and c); (ii) a sudden change of overall EGG signal amplitude (Figure 25e) (Roubeau *et al.*, 2009); (iii) an abrupt pitch change (Figure 25f), that was accompanied by an audible change of vocal timbre; and (iv) a sudden decrease of sound pressure level by 6 dB (not shown in figure). The change of EGG waveform happened over a period of ca. 0.028 seconds (7 glottal cycles).

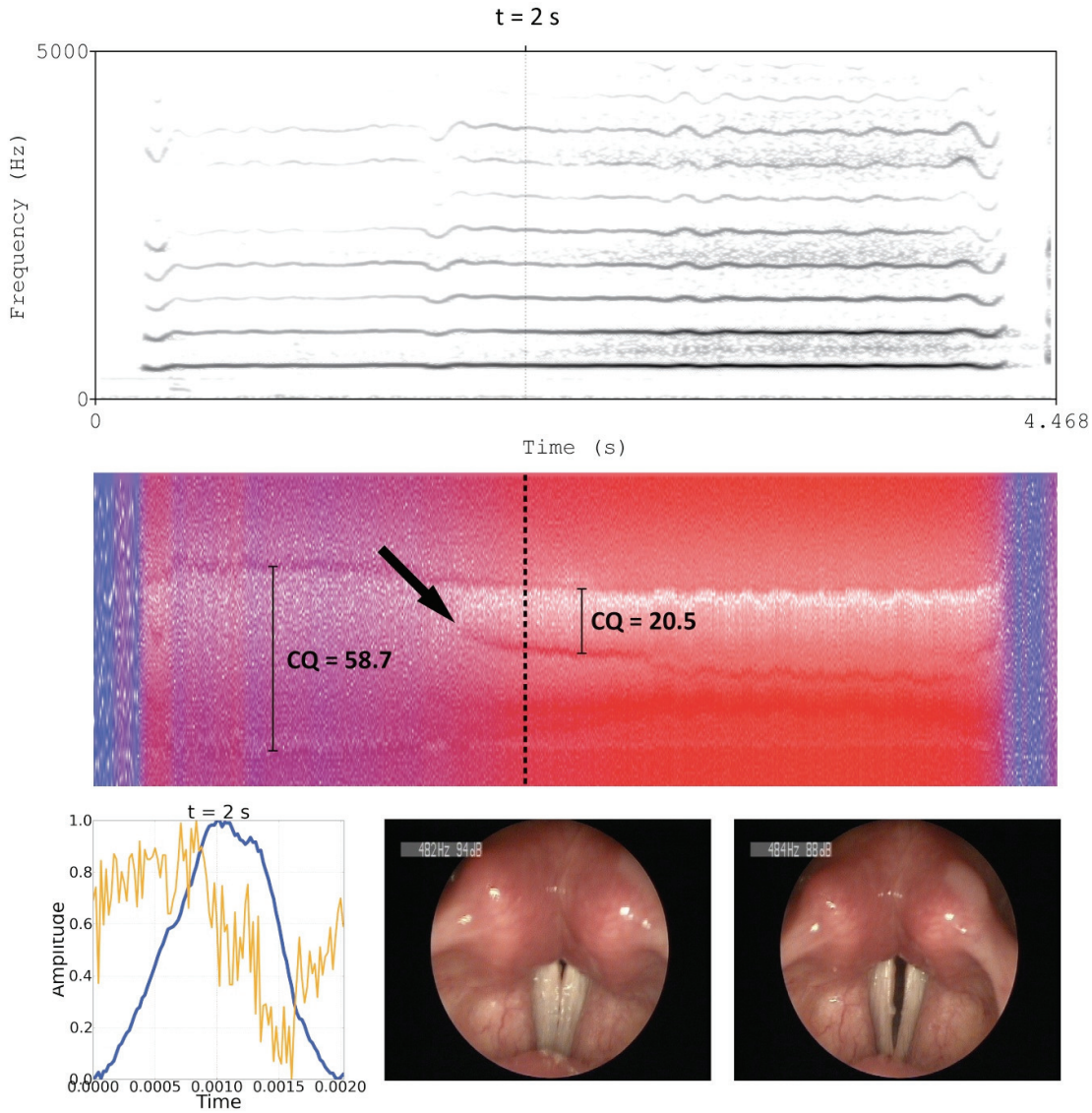


Figure 26 - Phonation at Bb4 (ca. 466 Hz) with increasing intensity by a woman with a vocal fold nodule on the right vocal fold. Top row: sound spectrogram of the microphone signal (FFT window length 0.05 s, dynamic range 50 dB); middle row: DEGG wavegram; bottom row, left: one glottal cycle, extracted at t = 2 s (EGG signal: blue; DEGG: orange); bottom row, middle and right: still images from videolaryngostroboscopic recording, taken at t = ca. 2 s at maximum closure and maximum vocal fold opening, respectively. The dark arrow indicates the establishment of vocal fold contact (see text for details). Note that the data presented in this graph is not included in manuscript III, as it has been assembled only recently.

A case of pathologic voice production is documented in Figure 26. Here, a new feature was introduced: the color coding of the wavegram. It is based on the overall amplitude of the EGG signal (calculated as the RMS energy of a sliding window with a length of 100 ms). Blue represents low amplitude (-20 dB in the documented case), and red represents high amplitude (-6 dB). Several features can be seen in this example:

- Based on the laryngoscopic evidence, vocal fold contact was established around t = 1.76 s, immediately after the overall amplitude of the EGG signal started to increase. At this point, an additional dark line (representing a peak in the DEGG waveform) develops in the wavegram (indicated by the black arrow).
- Inspection of the video data revealed that the "pseudo contact phase" seen in the wavegram from t = 0.2 s to t = 1.76 s (having a contact quotient of 0.59) was caused by contact of the vocal nodules, and possibly by a mucus bridge. Proper vocal fold contact, as seen after t = 1.76, had a contact quotient of ca. 0.20, increasing to ca. 0.27 in the course of the phonation.
- The fluctuation in the EGG waveform alignment in the wavegram, developing at ca t = 2.6 s, is phase shifted by 180° in relation to the fundamental frequency.
- The first derivative of the EGG waveform extracted at t = 2 s would not suggest any peak as being useful for detection of a vocal fold contacting event. Yet, in the context of the wavegram, gradual developments over time are made transparent.

3.2.3. Discussion

In the wavegram, the EGG waveform is treated "as is", and it's full informative content is retained. Thus, the wavegram constitutes an alternative to contact quotient monitoring, since it provides additional information, and does not depend on any arbitrary threshold criterion (Herbst and Ternström, 2006). Our observations suggest that this can provide novel insights into details of vocal fold behavior that are easily overlooked in other analysis techniques.

For instance, the wavegram reveals that vocal fold contacting and de-contacting "events" are more complex than commonly assumed. Rather than a single incident, vocal fold contacting and de-contacting should be considered a sequence of events, taking place over a certain period of time. The analyses presented here indicate that this sequence of events can change with fundamental frequency, loudness and register.

Vocal fold vibration is a complex phenomenon: The vocal folds do not merely open and close as two monolithic blocks. Rather, phase differences along the anterior-posterior and/or inferior/superior dimension are seen, as well as various other features, caused by physiological or pathological changes introduced into the vocal fold shape. In the EGG signal, the complex vibratory motion (taking place in three dimensions) is captured as one single time-varying value (relative vocal fold contact area). The data presented in this dissertation suggests however, that the EGG signal conveys more physiological information on vocal fold vibration than what is offered by more traditional representations. The wavegram technique promises to provide a method to visualize, further explore and understand this "hidden" information.

The rugged appearance of the wavegram displayed in Figure 26 is different from what is seen in Figure 24 and Figure 25. In particular, (a) multiple interspersed positive and negative maxima of the DEGG waveform (within each glottal cycle) and (b) cycle-to-cycle variations of these DEGG peaks might hint at abnormal vocal fold vibration both on an intra-cycle and an

inter-cycle level. This example suggests that the EGG wavegram might be useful for analysis of pathologic voices, at least when phonation is periodic and the fundamental frequency can be calculated. The documented case also suggests that the new method is also conceivably suitable for screening of voice disorders. More research is necessary to assess the true potential of the wavegram for these purposes.

3.2.4. Conclusions

The wavegram technique provides a new and potentially powerful method for displaying entire electroglottographic signals, or parts thereof. Much like in the spectrogram, information on vibratory behavior developing in time is compacted into one single graph, hence allowing to observe changes of vocal fold dynamics. Nevertheless, waveform details of individual glottal cycles (and their gradual development over time) are preserved, thus providing a useful tool to quickly gain more physiologic insights into longer bouts of phonation than what is visible in a simple amplitude graph.

The wavegram reveals changes of vocal fold contact phase, as well as of phenomena that cannot easily be seen with other methods for displaying electroglottographic signals/waveforms. Those phenomena, which develop over time, are expected to be related to physiologic behavior of vocal fold vibration, the nature of which has not yet been fully understood and explored.

3.3. Paper II – Four phonation types

3.3.1. Introduction

This part of the dissertation, based on manuscript **II**, investigated a classically-trained baritone who demonstrated the ability to produce four distinct voice qualities, which he considered to be the 'building blocks' of his singing technique. He called these four qualities: Type A) *' Naïve singer's falsetto';* Type B) *'Counter-tenor falsetto'*; Type C) *'Lyrical style'*; and Type D) *'Full chest'*. Acoustic recordings of these four phonation types are demonstrated in the audio file accompanying manuscript **II** (http://dx.doi.org/10.1121/1.3057860). Based on his self-perception, the subject claimed that these four voice qualities are created with four different laryngeal configurations. The specific goal of this study was to investigate these phonations and the laryngeal adjustments in these four phonation types. The more general goal was to establish a better understanding of the laryngeal adjustment strategies that are used to control the voice quality in singing.

3.3.2. Methods

The aforementioned four voice qualities were produced as sustained phonations ("endoscopic" vowel /i/) at the fundamental frequency of ca. 294 Hz (tone D4, at the baritone's *zona di passaggio*[17]). The adjustment and the vibration of the vocal folds were observed with laryngeal videostroboscopy (Baken and Orlikoff, 2000; Bless *et al.*, 1987) and

[17] The *zona di passaggio* is a pitch region where phonation in both chest and falsetto register is possible.

videokymography (Baken and Orlikoff, 2000; Švec and Schutte, 1996). Phonation was also documented with acoustic and electroglottographic recordings.

The videokymographic closed quotient (CQ_{VKG}) was determined from selected representative videokymographic recordings, using the formula $CQ_{VKG} = t_c/T_0$. Here, t_c is the duration of the closed phase and T_0 is the duration of the vibratory cycle. The durations were measured by manually counting the number of pixels for open and closed phase. For each of the 15 samples of stable phonation, the CQ_{VKG} was calculated from four consecutive glottal cycles at two or three (in short or long recordings, respectively) different instances equally distributed over the duration of the phonatory sample. CQ_{VKG} readings were then averaged per phonation type.

The alpha ratio was calculated as the ratio of the high-band energy (1000 – 5000 Hz) and the low-band energy (up to 1000Hz) of the acoustic signal, expressed in dB (Löfqvist and Mandersson, 1987). Typical alpha ratio values are negative numbers, which increase (i.e become less negative) as the spectrum of the signal becomes flatter (i.e. the signal has stronger high-frequency components).

The electroglottographic contact quotient (CQ_{EGG}) was calculated by a criterion level method with a threshold level of 25 % as has been used, e.g., by Orlikoff (1991) or Herbst & Ternström (2006). Since the electroglottograph used for data acquisition had no automatic gain controller, it was also possible to evaluate the strength of the EGG signal, which was quantified by computing the time-varying RMS value of the signal, using a window duration of 125 ms. The values were expressed on a dB scale with an arbitrary reference value.

3.3.3. Results

Examples of the videostroboscopic recordings of the four phonation types are given in the video file that was published online together with manuscript (**II**), which can be found at http://dx.doi.org/10.1121/1.3057860. Figure 27 reveals distinct adjustments of the laryngeal structures and vibratory features of the vocal folds for each phonation type. The most distinct differences were seen

(1) at the posterior, cartilaginous part of glottis which was varying between slightly open (Type A and C) and closed (Type B and D);

(2) at the vocal processes of the arytenoid cartilages (marked by arrows in the stroboscopic images of Figure 27, row 1) which were in some cases vibrating with the vocal folds (Type A and C) and in other cases pressed together and not-vibrating (Type D; in phonation type B, the vocal processes were intermittently pressed together or vibrating with the vocal folds); and

(3) in the mucosal waves on the vocal folds the extent of which varied from long (Type D) to very short (Type A and B).

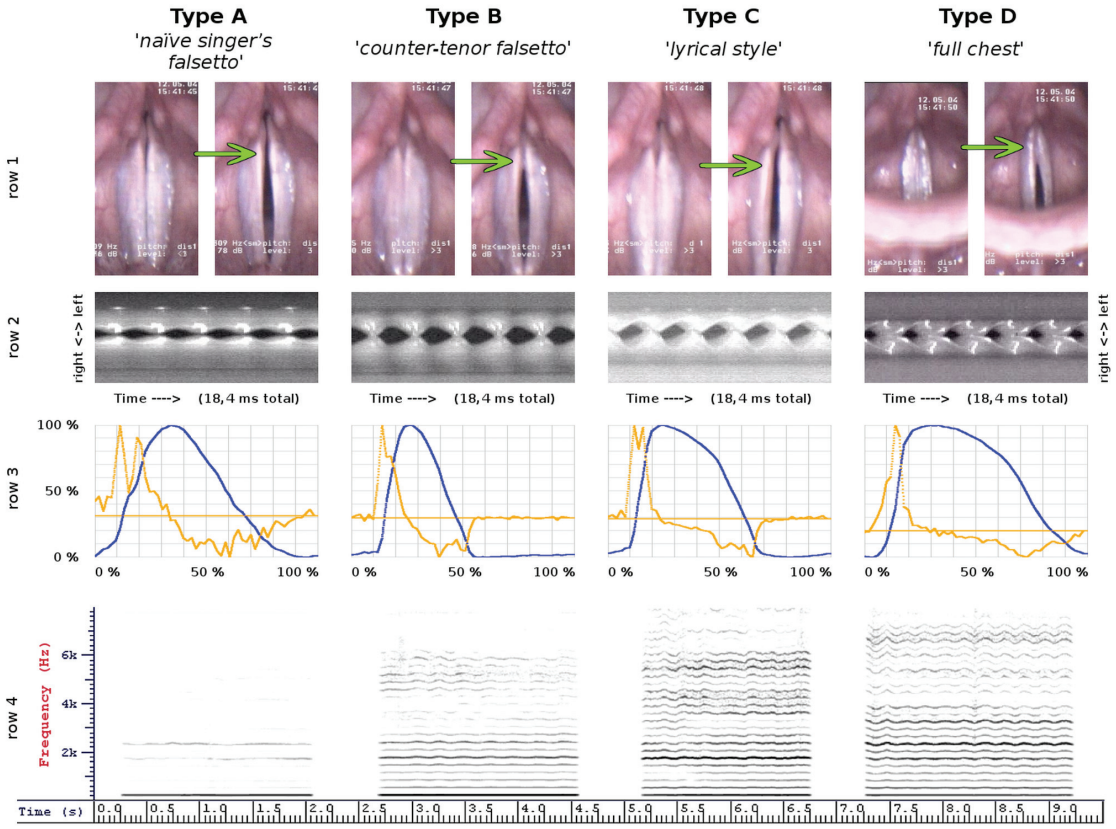


Figure 27 - Images and signals documenting the vocal fold vibration in the four phonation types A-D. From top to bottom: [Row 1] pairs of videostroboscopic images at the phases of maximum vocal fold contact and maximum glottal opening. The arrows point at the position of the vocal processes of the arytenoids cartilages which are apart in Types A and C and closed in Types B and D; [Row 2] typical videokymographic images at the place of maximal vibration amplitude of the vocal folds; [Row 3] typical EGG signals (blue) and their first derivative (gold) for phonation types A through D, normalized both in amplitude and time. The x-axis represents normalized glottal cycle duration, and the y-axis shows EGG and DEGG signal amplitude, locally normalized per glottal cycle. The overall EGG signal amplitude for phonation type A was ca. 8 – 10 dB weaker than that of the other phonations; [Row 4] spectrogram of sustained notes produced with phonation types A, B, C and D consecutively.

The videokymographic images revealed distinct differences in the vibratory pattern of the vocal folds among the four phonation types (Figure 27, row 2). Spectral analysis showed an increasing energy content in high frequency partials from type A to D. Please refer to manuscript **II** for a detailed description of the results.

3.3.4. Discussion

The collected data suggests that the four phonation types were produced with four different laryngeal adjustments. The main distinctive features were the configuration of the posterior glottis and the position of the vocal processes. It may thus be appropriate to call the phonation types A and B "aBducted falsetto" and "aDducted falsetto" (referred to as "open-chink falsetto" and "closed-chink falsetto" by Rubin and Hirt (1960), whereas the types C and D can be seen as "aBducted chest" and "aDducted chest" registers, respectively.

The present results suggest that in singing, the adduction of the vocal processes is actively used as a parameter in voice quality control (recall chapter 1.6.2), which is independent and separate from that of the (chest-falsetto) register control by the vocalis muscle (recall chapter 1.6.3.1). While the posterior glottal adduction has been recognized as an independent parameter in voice production (e.g. Titze, 2000), the use of this parameter separately from the chest-falsetto control in singing has, to the author's knowledge, not yet been documented. An independent control and flexibility of the adductory muscles appears to be an important factor in allowing the production of different singing qualities.

3.3.5. Conclusions

The results of this study showed that the four phonation types were acoustically distinct, and that they were indeed produced with different laryngeal settings. These settings could be explained by the independent manipulation of mainly two laryngeal parameters: 1) *the thickening of the vocal folds* and 2) *the adduction of the posterior glottis*. These two physiologic parameters represent two physiologically distinct types of glottal adduction: membranous adduction (adjustable by thyroarytenoid muscle) and cartilaginous adduction (adjustable by cricoarytenoid and interarytenoid muscles). The two types of glottal adduction should be separated from each other when studying different voice qualities in singing. Examination of a larger pool of subjects is required, in order to find out whether these findings can be generalized to trained and untrained singers of both sexes.

3.4. Paper III - Laryngeal control of sound source properties

3.4.1. Methods

The four singing voice qualities described in manuscript **II** were taken as a basis for the investigations in this part of the dissertation. They were labeled as 'aBducted falsetto' (FaB), 'aDducted falsetto' (FaD), 'aBducted chest' (CaB) and 'aDducted chest' (CaD). 13 subjects (six female and seven male singers and non-singers) received a 30 minute training session, in which they were presented with the four phonation types as demonstrated by an instructor (author C.T.H., who was the subject in the previous pilot study, see manuscript **II**). The singers were asked to vocally imitate the instructor until a consensus was reached between the instructor and the individual subjects that they did achieve the targeted phonation type.

During the training session, each subject's *zona di passaggio* was established as the range of pitches at which the target notes of all four phonation types could be produced. In order to make the desired registration (chest or falsetto) easier, the target notes were reached by singing a descending (for falsetto) or ascending (for chest) scale of five notes. The subjects were asked not to 'blend or mix the registers'. The vowel /i/ which allows examination through rigid laryngoscopy, was used for all phonatory tasks. After the training session, the four targeted phonation types were recorded during simultaneous capture of acoustic and electroglottographic data and laryngeal imaging (videostrobolaryngoscopy and VKG).

The cartilaginous adduction was evaluated by measuring the posterior glottal chink area (A_{PGC}) at maximum glottal closure, specified in pixels. The A_{PGC} data was averaged per subject over the four phonation types, and each individual A_{PGC} value was divided by this average as a means of intra-subject data normalization. The membranous glottal adduction was estimated indirectly by measuring the videokymographic closed quotient CQ_{VKG} at the middle of the membranous portion of the glottis (Herbst and Ternström, 2006; Qiu *et al.*, 2003). For both

CQ_{VKG} and glottal area data, the difference among all phonation types was evaluated statistically.

Please refer to manuscript **III** for a detailed description of the protocol, subject demographic data, recording procedures, equipment, and data analysis.

3.4.2. Results

All the subjects had a greater posterior glottal chink and thus a less adducted posterior glottis in the two aBducted (breathy) phonation types ('aBducted falsetto' and 'aBducted chest') than in the two aDducted (non-breathy) phonation types ('aDducted falsetto' and 'aDducted chest') – see manuscript **III**, and Figure 27, row 1, for comparable results.

The differences between chest and falsetto phonations were documented by analysis of the duration of glottal closure based on videokymographic (VKG) data - see manuscript **III**, and Figure 27, row 2, for comparable results. In overall, the chest register phonations had a general tendency to have a larger glottal closure duration (and hence a larger CQ_{VKG}) than their falsetto counterpart. The duration of glottal closure per glottal cycle also varied within both the chest and falsetto register when changing the posterior adduction. The closed quotient CQ_{VKG} rose when changing from aBducted falsetto to aDducted falsetto, as well as when changing from aBducted chest to aDducted chest. As an unexpected result, in five out of 13 subjects the CQ_{VKG} values reached similar or even larger values for the aDducted falsetto (FaD) than for the aBducted chest (CaB).

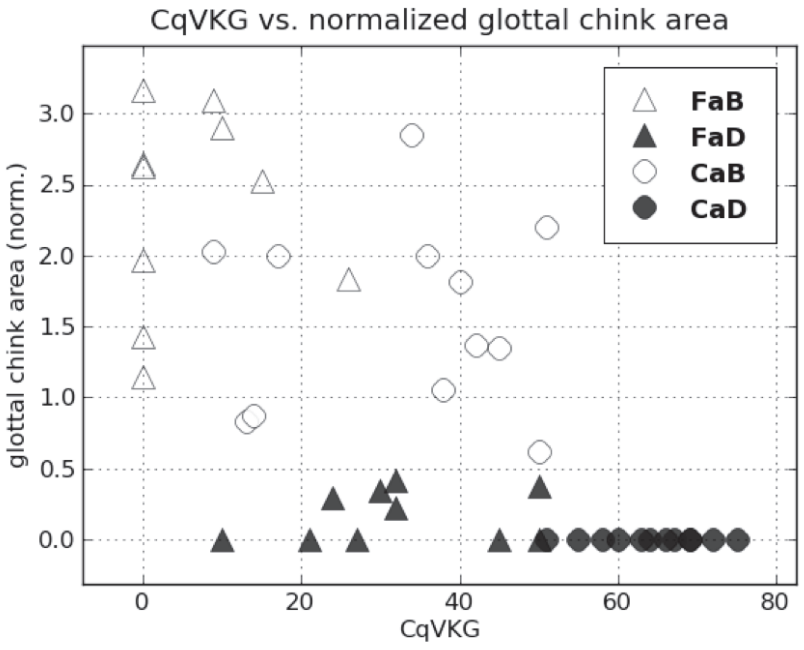


Figure 28: Graphical illustration of the relation of videokymographic closed quotient to glottal chink area for all subjects. Note that the aBducted falsetto phonations (FaB) are found in the upper left corner (signifying a small closed quotient and a large glottal chink area), and the aDducted chest (CaD) phonations are located in the lower right corner (representing a fully adducted posterior glottis, and a large closed quotient).

A comparison of videokymographic closed quotient CQ_{VKG} and glottal chink area data for each individual phonation type and each subject is shown in Figure 28. Here, all the aDducted chest phonations are found in the lower right corner of the graph (having a high CQ_{VKG} and no posterior glottal chink, i.e. fully adducted posterior glottis). On the other hand, the aBducted falsetto phonations are located in the upper left corner of the graph, since they were generally produced with a large posterior glottal gap and a small CQ_{VKG}. The aDducted falsetto phonations and the aBducted chest phonations occupy the central area of the graph. They can be distinguished from each other by the difference in posterior glottal chink area (the aBducted chest phonations generally showing a larger posterior glottal chink). The adductory changes needed for transitions between the phonation types for all subjects are summarized in Table 1 (see manuscript **III** for further details).

Phonation type transition	**Cartilaginous adduction**	**Membranous medialization**
FaB -> FaD	**++**	**++**
FaB -> CaB	**N/A**	**++**
FaB -> CaD	**++**	**++**
FaD -> CaB	**--**	**N/A**
FaD -> CaD	**+**	**++**
CaB -> CaD	**++**	**++**

Table 1: Summary of laryngeal adjustments for transitions between phonation types. ++, --: statistically significant changes; +: statistically insignificant changes, but slight trend observed in some subjects; N/A: no general trend observed, no statistical significance

3.4.3. Discussion

The results for this study showed distinct laryngeal adjustments for the four phonation types in most of the subjects. The aBducted phonations had a larger glottal chink area than their respective aDducted counterparts, and the chest phonations exhibited a longer closed phase (i.e. larger CQ_{VKG},) than the respective falsetto phonations.

It has been generally accepted that the duration of the closed phase is greater in chest register than in falsetto (Choi *et al.*, 1993; Henrich *et al.*, 2005; Hirano, 1981; Roubeau *et al.*, 2009; Salomão, 2008; Vilkman *et al.*, 1995). Our data, based on CQ_{VKG} calculations, indicate that this assumption is only valid for the cases where the posterior adduction in falsetto is the same or reduced, as compared to the chest register. When the chest register is produced with an aBducted posterior glottis, the CQ_{VKG} values may sometimes reach equal or smaller values than in aDducted falsetto phonations. This hitherto unreported finding implies that the closed quotient (measured at the place of maximum vibration amplitude) should not be used as a sole indicator of voice registers in singing.

A closer look at the mechanics of vocal fold posturing suggests that the *cartilaginous adduction*, i.e. the adduction of the arytenoids, has two effects: (a) the adduction of the posterior, cartilaginous part of the glottis; and (b) the adduction of the membranous portion of the vocal folds. The membranous vocal fold portion is, however, also adducted by another adjustment of the vocal folds, which we call *membranous medialization through vocal fold*

bulging (short: *membranous medialization*). This adjustment is caused by an active increase of the volume of the membranous portion of the vocal folds through the activity and bulging of the thyroarytenoid muscle. The membranous adduction can thus be targeted separately through (a) the membranous medialization, without largely affecting the configuration of the cartilaginous glottis (Nasri *et al.*, 1994); and (b) as a by-product of cartilaginous adduction (see the discussion above). Schematically, the adduction and glottal configurations of the four phonation types, along with the expected roles of the internal laryngeal muscles, are displayed in Figure 29.

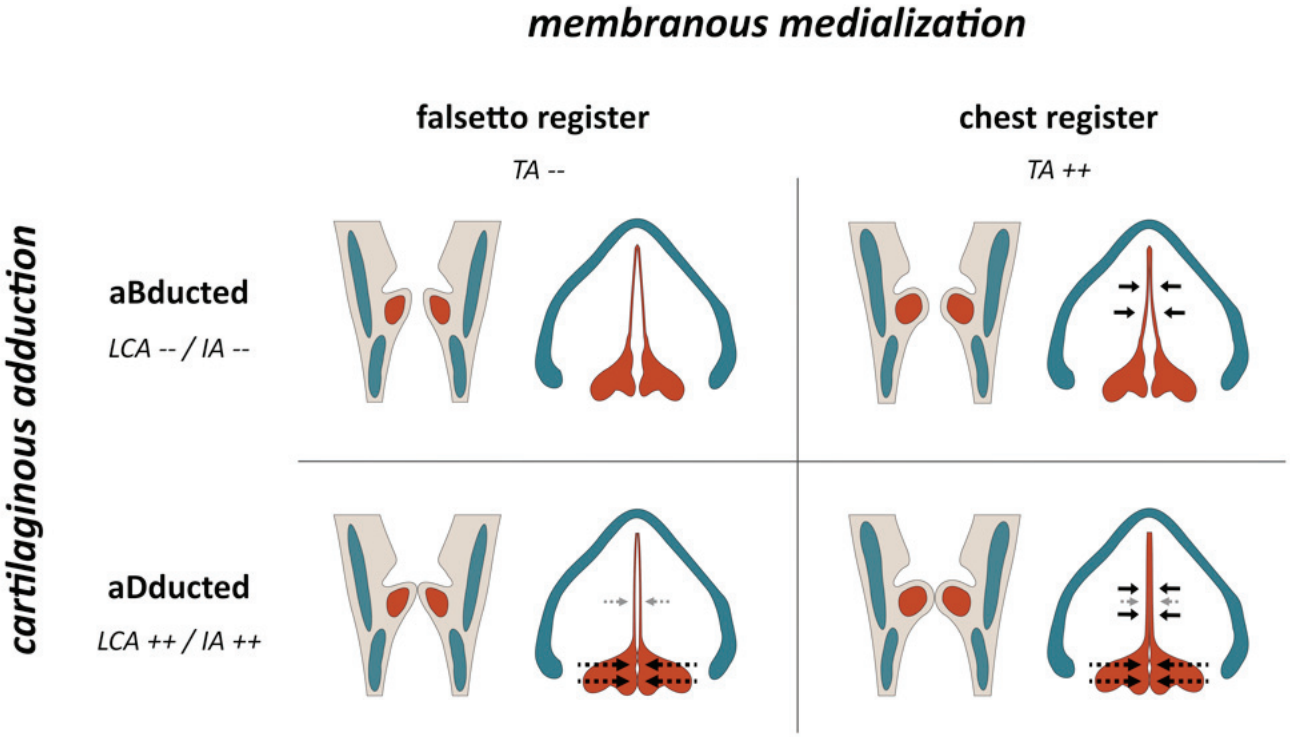


Figure 29: Schematic illustration of the effect of cartilaginous adduction and membranous medialization through vocal fold bulging in singing. For each adduction type, two schematic graphs are shown: top view of vocal folds, arytenoids and thyroid cartilage (left); sagittal view of larynx, with schematic drawings of thyroid cartilage, cricoid cartilage, and thyroarytenoid muscle (right). The arrows indicate the primary changes in the vocal fold position for each case. The expected theoretical contributions of the internal laryngeal muscles (TA, LCA and IA) are also indicated.

The designed phonatory exercises allowed the subjects (both trained and untrained) to separately manipulate the cartilaginous adduction (by changing the degree of posterior vocal fold adduction) and membranous medialization (by changing the register of phonation). These results suggest that the four targeted phonation types produced by the subjects were created by varying the degrees of cartilaginous adduction and membranous medialization – see Figure 30.

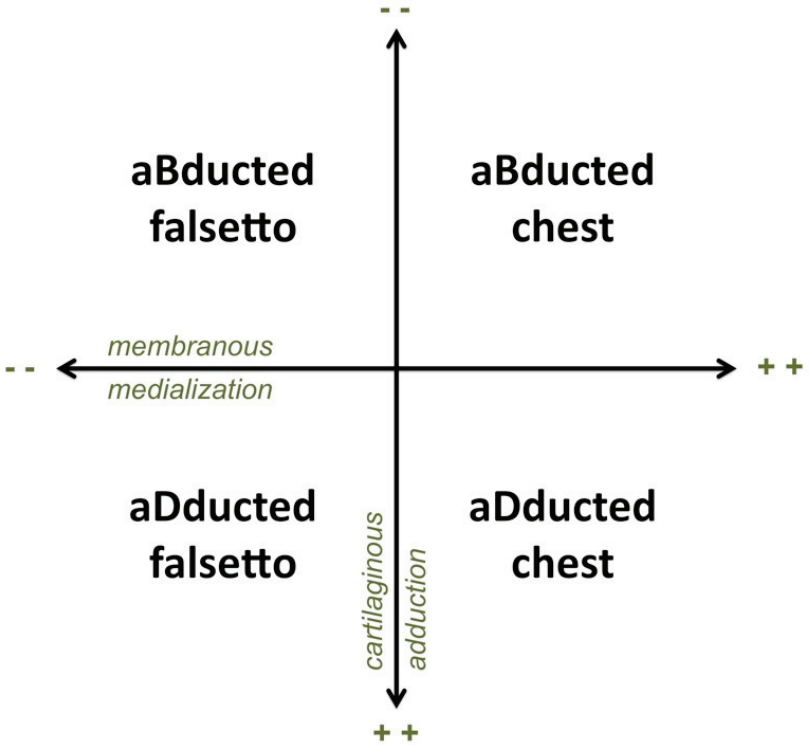


Figure 30: Relationship between the two types of adduction (membranous medialization and cartilaginous adduction) and the four phonation types described here (aBducted falsetto, aDducted falsetto, aBducted chest, aDducted chest). This pedagogical model is not limited to distinct (and possibly extreme) phonation types. Rather, it encourages gradual, fine-controlled adjustments along the two major dimensions, in order to increase timbral variability for enhanced artistic expression.

3.4.3.1. Two degrees of freedom at the sound source – a pedagogical model

Since cartilaginous adduction and membranous medialization can be controlled separately by both trained and untrained singers, these two physiological parameters can be displayed in a two-dimensional plane in order to create a pedagogical model for sound quality adjustments made on the laryngeal level – see Figure 30. This model consists of four quadrants, representing the four sound qualities described above. Different singing styles can be mapped onto the model, showing their main mechanism of production, such as[18]:

- **aBducted falsetto:** e.g. untrained female classical singing (high tessitura); "naïve" male falsetto; "lighter" registration (both male and female) in pop and jazz.
- **aDducted falsetto:** e.g. trained female classical; counter-tenor;
- **aBducted chest:** e.g. untrained naïve singing (lower range); "lighter" registration (both male and female) in pop and jazz.
- **aDducted chest:** e.g. trained male classical; belting; rock; soul;

[18] This is a non-exhaustive list, intended solely to provide some basic orientation (further research is needed to assess and objectify the assumptions made here). Please note that some of the singing styles could also make use of other phonation types.

The model is in no ways restricted to the production of four distinct phonation types. On the contrary, gradual adjustments along the two major axes (cartilaginous adduction and membranous medialization) are possible. These fine-controlled muscular adjustments allow experienced singers to create a large variety of sound timbres at the glottal level, increasing their expressional freedom within the aesthetical boundary conditions of their chosen singing style. Also, the model can provide pedagogical strategies, suggesting physiological adjustments to successfully create desired singing styles.

3.4.3.2. A model case from the singing studio

A typical example for the application of the described model of two types of glottal adduction would be the mid-range of female classical singers. We postulate that the desired quality above the *primo passaggio* (around pitch D4, ca. 295 Hz[19]) would have to be aDducted falsetto, since (a) in classical singing (as opposed to other singing styles, such as belting), chest phonation is usually restricted to the lower range; and (b) breathy phonation, created by aBducted falsetto, is both acoustically inefficient, and outside the aesthetical boundaries of classical singing. Nevertheless, untrained female singers have a tendency to produce sounds above the primo passaggio in either aDducted chest register, or in aBducted falsetto. In other words, they either "push" the chest register up (usually at the risk of an abrupt register transition at the upper end of the chest register range), or they produce soft breathy sounds, in both cases severely limiting the dynamic range and timbral variability of the voice at those particular pitches.

The difficulty of switching from aDducted chest to aDducted falsetto can be explained physiologically: when changing from chest to falsetto register, the TA muscle relaxes and thus membranous medialization is reduced. The resulting decrease of overall vocal fold adduction could be counteracted by a slight increase of cartilaginous adduction via the LCA and IA muscles, in order to keep the timbral change at a minimum. Such a maneuver (decreasing TA whilst increasing LCA and IA activity, respectively) is, however, both ambivalent and complex: Since the three involved muscles are all innervated by the recurrent laryngeal nerve, the required level of fine control might be hard to reach by some singers.

A valid pedagogical strategy would try to establish the aDducted falsetto at higher pitches (around Bb4 to F5), via calling (but not shouting) at vowel /u/, or using "primal sounds" (Chapman, 2006). This sound quality could then be applied to other vowels and pitches all the way down to the primo passaggio. When attempting to achieve a successful transition from (aDducted) chest to aDducted falsetto on ascending scales, there is the danger that the singer "gets stuck" in the chest register. In order to avoid this, it is advisable that ascending scales are only sung when:

(a) the student reliably succeeded to produce aDducted falsetto in the primo passaggio on descending scales; and

(b) s/he managed the falsetto-chest register transition at the primo passaggio without abrupt timbral changes.

[19] Depending on anatomy and *Fach*, but hardly exceeding the range of C4 to F4.

3.4.4. Conclusions

Both singers and non-singers (of both sexes) can independently control the *cartilaginous adduction* and *membranous medialization* (by means of adduction of the arytenoid cartilages and the bulging of the vocal folds respectively). Independent control over cartilaginous adduction and membranous medialization is particularly important for the singing voice, helping the experienced singer to fine-tune the characteristics of the sound source. In this respect, the exercises described here can be useful for singers with limited flexibility who experience problems with producing different timbres.

The gathered data also showed that the videokymographically derived closed quotient should not be considered to be the sole indicator of the voice register in singing, since it can in some subjects achieve larger values in 'aDducted falsetto' than in 'aBducted chest' phonations. This is attributed to the fact that membranous adduction (and thus the closed quotient) is influenced by both membranous medialization and cartilaginous adduction.

The two physiological parameters described here can be mapped onto a two-dimensional plane in order to create a pedagogical model for sound quality adjustments made on the laryngeal level. This model promises to be useful in better understanding different singing styles, and for applications in the voice studio.

4. Overall conclusion

In recapitulation, the aims of this dissertation were:

1. to develop a new method for visualization and analysis of the electroglottographic signal, i.e. a physiological correlate of vocal fold motion; This method should allow to display longer EGG signals in one graph, whilst retaining the appearance of each individual glottal cycle's waveform, as found in the original EGG signal.
2. to document physiological means of sound source control via adjustments of the vocal folds, by investigating the relation between voice registers and (posterior) glottal adduction.

The first of these goals was reached by introducing **EGG wavegrams**. This new technique, as presented here, enables us to better tap the rich potential of the EGG waveform. Wavegrams allow assessment and analysis of the EGG waveforms more comprehensively, and promise to enhance our understanding of the EGG signal and vocal fold vibration. Wavegrams might in the future become a useful tool for analysis of the electroglottographic signal, which holds a great potential for singing pedagogy (Herbst *et al.*, 2010; Howard *et al.*, 2004; Nair, 1999; Rossiter *et al.*, 1996), and possibly for detecting voice disorders (Baken and Orlikoff, 2000; Smith and Childers, 1983) and in voice therapy (Baken and Orlikoff, 2000; Fourcin *et al.*, 1995).

The second goal of this thesis was achieved by recognizing and describing the separate adjustment of **cartilaginous adduction** and **membranous medialization**, and their effects on the sound source via the chosen glottal configuration. This new understanding of these two physiological parameters has great promise for research, voice therapy and singing pedagogy. The model suggested here (see chapter 3.4.3.1) allows both singing teachers and voice therapists to better understand and diagnose their clients' voice production habits on a physiological level. This could eventually lead to novel treatment and pedagogical methods (or to better comprehension and application of existing ones), thus increasing the efficiency and effectiveness of therapy and teaching.

5. References

ANSI (**1960**). "USA Standard Acoustical Terminology (Including Mechanical Shock and Vibration)," Technical Report **S1.1-1960 (R1976)**.

Asher, V. A., Sasaki, C. T., and Gracco, C. (**1996**). "Laryngeal Physiology. Normal and Abnormal.," in *The Larynx*, edited by M. Fried (Mosby, St. Louis), pp. 45-54.

Baer, T. (**1975**). *Investigation of phonation using excised larynxes. Doctoral dissertation* (Massachusetts Institute of Technology, Cambridge, Mass).

Baer, T., Löfqvist, A., and McGarr, N. S. (**1983**). "Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques," J Acoust Soc Am **73**, 1304-1308.

Bailly, L., Henrich, N., and Pelorson, X. (**2010**). "Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling," J Acoust Soc Am **127**, 3212-3222.

Baken, R., and Isshiki, N. (**1977**). "Arytenoid displacement by simulated intrinsic muscle contraction," Fol Phon **29**, 206-216.

Baken, R. J. (**1992**). "Electroglottography," J Voice **6**, 98-110.

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice (2nd Edition)* (Singular Publishing, Thompson Learning).

Behnke, E. (**1886**). "The registers of the voice," Proceedings of the Musical Association **13th session**, 1-16.

Berry, D., Herzel, H., Titze, I., and Story, B. (**1996**). "Bifurcations in excised larynx experiments," J Voice **10**, 129-138.

Bless, D., Hirano, M., and Feder, R. (**1987**). "Videostroboscopic evaluation of the larynx," Ear, Nose and Throat Journal **66**.

Broad, D. (**1968**). "Kinematic considerations for evaluating laryngeal cartilage motions," Fol Phon **20**, 269-284.

Castellengo, M., Chuberre, B., and Henrich, N. (**2004**). "Is Voix Mixte, the Vocal Technique Used to Smoothe the Transition across the two Main Laryngeal Mechanisms, an Independent Mechanism? ," in *Proceedings of the International Symposium on Musical Acoustics (ISMA2004)* (Nara, Japan).

Chapman, J. (**2006**). *Singing and teaching singing. A holistic approach to classical voice* (Plural Publishing, San Diego, Oxford, Brisbane).

Childers, D., and Krishnamurthy, A. K. (**1985**). "A critical review of electroglottography," Crit Rev Biomed Eng **12**, 131-161.

Childers, D., Naik, J. M., Larar, J. N., Krishnamurthy, A. K., and Moore, G. P. (**1983**). "Electroglottography, speech, and ultra-high speed cinematography," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze (The Denver Center for the Performing Arts), pp. 203-220.

Childers, D. G., Hicks, D. M., Moore, G. P., and Alsaka, Y. A. (**1986**). "A model for vocal fold vibratory motion, contact area, and the electroglottogram," J Acoust Soc Am **80**, 1309-1320.

Choi, H. S., Berke, G. S., Ye, m., and Kreiman, J. (**1993**). "Function of the thyroarytenoid muscle in a canine laryngeal model," Ann Otol Rhin Laryngol**102**, 769-776.

Coffin, B. (**1980**). *Coffin's Overtones of Bel Canto* (The Scarecrow Press, Metuchen, N. J.).

Colton, R. H., and Conture, E. G. (**1990**). "Problems and pitfalls of electroglottography," J Voice **4**, 10-24.

Deliyski, D. D., and Hillman, R. E. (**2010**). "State of the art laryngeal imaging: research and clinical implications," Curr Opin Otol Head Neck Surg**18**, 147-152.

Doscher, B. M. (**1994**). *The functional unity of the singing voice* (The Scarecrow Press, Metuchen, NJ and London).

Doval, B., d'Alessandro, C., and Henrich, N. (**2006**). "The Spectrum of Glottal Flow Models," Acta Acustica united with Acustica **92**, 1026-1046.

Echternach, M., Sundberg, J., Arndt, S., Markl, M., Schumacher, M., and Richter, B. (**2010**). "Vocal tract in female registers--a dynamic real-time MRI study," J Voice **24**, 133-139.

Echternach, M., Sundberg, J., Baumann, T., Markl, M., and Richter, B. (**2011a**). "Vocal tract area functions and formant frequencies in opera tenors' modal and falsetto registers," J Acoust Soc Am **129**, 3955-3963.

Echternach, M., Traser, L., Markl, M., and Richter, B. (**2011b**). "Vocal tract configurations in male alto register functions," J Voice **25**, 670-677.

Fabre, P. (**1957**). "Un procédé électrique percuntané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence; premiers

résultats (A non-invasive electric method for measuring glottal closure during phonation: high frequency glottography; first results)," Bull. Acad. Nat. Med. **141**, 66-69.

Fant, G. (**1960**). *Acoustic theory of speech production* (Mouton and Co., 's-Gravenhage).

Fant, G. (**1979**). "Glottal source and waveform analysis," STL-QPSR **1/1979**, 85-107.

Felsner, M. (**2008**). *Operatica: Annäherungen an die Welt der Oper* (Königshausen & Neumann, Würzburg).

Flanagan, J. (**1958**). "Some properties of the glottal sound source," J Speech Hearing Research **1**, 99-116.

Fourcin, A., and Abberton, E. (**1971**). "First application of a new laryngograph," Medical and Biological Illustration **21**, 172-182.

Fourcin, A., Abberton, E., Miller, D., and Howells, D. (**1995**). "Laryngograph: speech pattern element tools for therapy, training and assessment," European Journal of Disorders of Communication **30**, 101-115.

Fried, M. P., Meller, S. M., and Rinaldo, A. (**2009**). "Adult laryngeal anatomy," in *The Larynx*, edited by M. P. Fried, and A. Ferlito (Plural Publishing, San Diego, CA).

Friedrich, G., and Dejonckere, P. H. (**2005**). "Das Stimmdiagnostik–Protokoll der European Laryngological Society (ELS) - erste Erfahrungen im Rahmen einer Multizenterstudie," Laryngo-, Rhino-, Otologie **84**, 744-752.

Fuks, L., Hammarberg, B., and Sundberg, J. (**1998**). "A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences," Royal Institute of Technology - Speech, Music and Hearing Quarterly Progress and Status Report (Stockholm) **3/1998**, 49-59.

Garcia, M. (**1847a**). *Mémoire sur la voix humaine présenté à l'Académie des Sciences en 1840* (Imprimerie d'E. Duverger, Paris).

Garcia, M. (**1847b**). *Traité complet de l'art du chant* (Schott Paris).

Garcia, M. (**1894**). *Hints on singing* (Ascherberg, Hopwood and Crew, London).

Henrich, N. (**2006**). "Mirroring the voice from Garcia to the present day: Some insights into singing voice registers," Log Phon Vocol **31**, 3-14.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J Acoust Soc Am **115**, 1321-1332.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2005**). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency," J Acoust Soc Am **117**, 1417-1430.

Henrich, N., Kiek, M., Smith, J., and Wolfe, J. (**2007**). "Resonance strategies used in Bulgarian women's singing style: a pilot study," Log Phon Vocol **32**, 171-177.

Herbst, C., and Ternström, S. (**2006**). "A Comparison of Different Methods to Measure the EGG Contact Quotient," Log Phon Vocol **31**, 126-138.

Herbst, C. T., Howard, D. M., and Schlömicher-Thier, J. (**2010**). "Using Electroglottographic Real-Time Feedback to Control Posterior Glottal Adduction during Phonation," J Voice **24**, 72-85.

Herbst, C. T., Qiu, Q., Schutte, H. K., and Švec, J. G. (**2011**). "Membranous and cartilaginous vocal fold adduction in singing," J Acoust Soc Am **129**, 2253-2262.

Hertz, C. M., Lindstrom, M., and Sonesson, B. (**1970**). "Ultrasonic recording of the vibrating vocal folds: a preliminary report," Acta Otolaryngol (Stockh) **69**, 223-230.

Herzel, H., and Reuter, R. (**1997**). "Whistle register and biphonation in a child's voice," Fol Phon et Logopaedica **49**, 216-224.

Hess, M. M., and Ludwigs, M. (**2000**). "Strobophotoglottographic transillumination as a method for the analysis of vocal fold vibration patterns," J Voice **14**, 255-271.

Higgins, M. B., and Schulte, L. (**2002**). "Gender differences in vocal fold contact computed from electroglottographic signals: the influence of measurement criteria," J Acoust Soc Am **111**, 1865-1871.

Hirano, M. (**1974**). "Morphological Structure of the Vocal Cord as a Vibrator and its Variations," Fol Phon **1974**, 89-94.

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

Hirano, M., Kakita, Y., Kawasaki, H., Gould, W. J., and Lambiase, A. (**1981**). "Data from high-speed motion picture studies," in *Vocal Fold Physiology*, edited by K. N. Stevens, and M. Hirano (University of Tokyo Press, Tokyo), pp. 85-93.

Holmberg, E., Hillman, R., and Perkell, J. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and

female speakers in soft, normal, and loud voice," J Acoust Soc Am **84**, 511-529.

Holmer, N. G., Kitzing, P., and Lindstrom, K. (**1973**). "Echo glottography. Ultrasonic recording of vocal fold vibrations in preparations of human larynges," Acta Otolaryngol (Stockh) **75**, 454-563.

Howard, D. M. (**1995**). "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," J Voice **9**, 163-172.

Howard, D. M. (**1998**). "Instrumental voice measurement: uses and limitations," in *The voice clinic handbook*, edited by T. Harris, S. Harris, J. S. Rubin, and D. M. Howard (Whurr Publishers, London), pp. 323-382.

Howard, D. M., Lindsey, G. A., and Allen, B. (**1990**). "Toward the Quantification of Vocal Efficiency," J Voice **4**, 205-212.

Howard, D. M., and Murphy, D. T. (**2007**). *Voice Science, Acoustics, and Recording* (Plural Publishing, San Diego, CA).

Howard, D. M., Welch, G. F., Brereton, J., Himonides, E., DeCosta, M., Williams, J., and Howard, A. W. (**2004**). "WinSingad: A real-time display for the singing studio," Log Phon Vocol **29**, 135-144.

Hunter, E. J., Titze, I. R., and Alipour, F. (**2004**). "A three-dimensional model of vocal fold abduction/adduction," J Acoust Soc Am **115**, 1747-1759.

Husson, R. (**1950**). *Ètude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée. (Thesis)* (Paris).

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," The Bell System Technical Journal **51**, 1233-1268.

Jiang, J., Yumoto, E., Lin, S. J., Kadota, Y., Kurokawa, H., and Hanson, D. G. (**1998**). "Quantitative measurement of mucosal wave by high-speed photography in excised larynges," Ann Otol Rhin Laryngol**107**, 98-103.

Jiang, J. J., Zhang, Y., Kelly, M. P., Bieging, E. T., and Hoffman, M. R. (**2008**). "An automatic method to quantify mucosal waves via videokymography," Laryngoscope **118**, 1504-1510.

Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Acoustics: tuning of vocal tract resonance by sopranos," Nature **427**, 116.

Kania, R. E., Hans, S., Hartl, D. M., Clement, P., Crevier-Buchman, L., and Brasnu, D. F. (**2004**). "Variability of electroglottographic glottal closed quotients: necessity of standardization to obtain normative values," Arch Otolaryngol Head Neck Surg **130**, 349-352.

Klatt, D., and Klatt, L. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J Acoust Soc Am **87**, 820-857.

Kob, M., Henrich, N., Herzel, H., Howard, D. M., Tokuda, I., and Wolfe, J. (**2011**). "Analysing and understanding the singing voice: Recent progress and open questions," Current Bioinformatics **6**, 362-374.

Kochis-Jennings, K. A., Finnegan, E., Hoffman, H. T., and Jaiswal, S. (**in press**). "Laryngeal Muscle Activity and Vocal Fold Adduction During Chest, Chestmix, Headmix, and Head Registers in Females " J Voice.

Ladefoged, P. (**2001**). *A course in phonetics* (Harcourt, Orlando, FL).

Letson, J. A., Jr., and Tatchell, R. (**1997**). "Arytenoid movement," in *Professional voice: the science and art of clinical care*, edited by R. T. Sataloff (Singular Publishing Group, San Diego, CA), pp. 131-145.

Löfqvist, A., and Mandersson, B. (**1987**). "Long-time average spectrum of speech and voice analysis," Fol Phon **39**, 221-229.

Mallampati, S. R. (**1996**). "Anesthetic principles of airway management," in *The larynx. A multidisciplinary approach*, edited by M. Fried (Mosby, St. Louis), pp. 127-134.

McClellan, J. H., Schafer, R. W., and Yoder, M. A. (**1998**). *DSP first: a multimedia approach* (Prentice-Hall, Upper Saddle River, NJ).

McCulloch, T. M., Van Daele, D., and Ciucci, M. R. (**2011**). "Otolaryngology head and neck surgery: an integrative view of the larynx.," Head Neck **33**, S46-53.

McGlashan, J., Sadolin, C., and Kjelin, H. (**2007**). "Can vocal effects such as distortion, growling, rattle and grunting be produced without traumatising the vocal folds?," in *7th Pan-European Voice Conference (PEVOC)* (Groningen, The Netherlands).

Mehta, D., and Hillman, R. (**2008**). "Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods.," Curr Opin Otolaryngol Head Neck Surg **16**, 211-215.

Miller, D. G. (**2000**). *Registers in singing: empirical and systematic studies in the theory of the singing voice. Doctoral dissertation* (University of Groningen, Groningen, the Netherlands).

Miller, D. G. (**2008**). *Resonance in Singing* (Inside View Press, Princeton, NJ).

Miller, D. G., and Schutte, H. K. (**1993**). "Physical definition of the "flageolet register"," J Voice **7**, 206-212.

Miller, D. G., and Schutte, H. K. (**2005**). "'Mixing' the Registers: Glottal Source or Vocal Tract?," Folia Phoniatr Logop **57**, 278-291.

Miller, D. G., Švec, J. G., and Schutte, H. K. (**2002**). "Measurement of characteristic leap interval between chest and falsetto registers," J Voice **16**, 8-19.

Miller, R. (**1986**). *Structure of Singing, The: System and Art in Vocal Technique* (Schirmer Books, New York).

Miller, R. L. (**1959**). "Nature of the vocal cord wave," J Acoust Soc Am **31**, 667-677.

Moore, P. (**1991**). "Voice: a historical perspective. A short history of laryngeal investigation," J Voice **5**, 166-281.

Mörner, M., Fransson, F., and Fant, G. (**1963**). "Voice register terminology and standard pitch," STL-QPSR **4/1963**, 17-23.

Nair, G. (**1999**). *Voice – tradition and technology. Using the spectrograph as a feedback tool in the voice studio* (Singular Publishing Group).

Nasri, S., Sercarz, J. A., Azizzadeh, B., Kreiman, J., and Berke, G. S. (**1994**). "Measurement of adductory force of individual laryngeal muscles in an in vivo canine model," Laryngoscope **104**, 1213-1218.

Nawka, T., and Wirth, G. (**2008**). *Stimmstörungen. Lehrbuch für Ärzte, Logopäden, Sprachheilpädagogen und Sprechwissenschaftler* (Deutscher Ärzte-Verlag, Köln).

Neubauer, J., Edgerton, M., and Herzel, H. (**2004**). "Nonlinear Phenomena in Contemporary Vocal Music," J Voice **18**, 1-12.

Orlikoff, R. F. (**1991**). "Assessment of the dynamics of vocal fold contact from the electroglottogram: data from normal male subjects," J Speech Hearing Res **34**, 1066-1072.

Qiu, Q., Schutte, H. K., Gu, L., and Yu, Q. (**2003**). "An automatic method to quantify the vibration properties of human vocal folds via videokymography," Fol Phon Logop **55**, 128-136.

Reid, C. L. (**2005**). "Voice Science: An Evaluation," Australian Voice **11**, 6-24.

Remmers, J. E., and Gautier, H. (**1972**). "Neural and mechanical mechanisms of feline purring," Resp Physiol **16**, 351-361.

Roads, C. (**1996**). "Spectrum Analysis," in *The computer music tutorial*, edited by C. Roads (The MIT Press, Cambridge, MA), pp. 533-612.

Roads, D., and Strawn, J. (**1998**). "Digital Audio Concepts," in *The Computer Music Tutorial*, edited by C. Roads (The MIT Press, Cambridge, MA).

Rossing, T. (**1990**). *The Science of Sound* (Addison-Wesley Publishing Company).

Rossiter, D., Howard, D. M., and DeCosta, M. (**1996**). "Voice development under training with and without the influence of real-time visually presented biofeedback," J Acoust Soc Am **99**, 3253-3256.

Rothenberg, M. (**1973**). "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," J Acoust Soc Am **53**, 1632-1645.

Rothenberg, M. (**1977**). "Measurement of airflow in speech," J Speech Hearing Res **20**, 155-176.

Rothenberg, M. (**1979**). "Some relations between glottal air flow and vocal fold contact area," in *Proceedings of the Conference on the Assessment of Vocal Pathology*, edited by C. L. Ludlow, and M. O. Hart (American Speech and Hearing Association), pp. 88-96.

Rothenberg, M. (**1981a**). "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology*, edited by K. N. Stevens, and M. Hirano (University of Tokyo Press, Tokyo), pp. 305-328.

Rothenberg, M. (**1981b**). "The voice source in singing," in *Research aspects on singing* (Royal Swedish Academy of Music, Stockholm), pp. 15-33.

Rothenberg, M., and Mahshie, J. J. (**1988**). "Monitoring vocal fold abduction through vocal fold contact area," J Speech Hearing Res **31**, 338-351.

Roubeau, B., Henrich, N., and Castellengo, M. (**2009**). "Laryngeal vibratory mechanisms: the notion of vocal register revisited," J Voice **23**, 425-438.

Rubin, H. J., and Hirt, C. C. (**1960**). "The falsetto. A high-speech cinematographic study," Laryngoscope **70**, 1305-1324.

Sadolin, C. (**2008**). *Complete Vocal Technique* (Shout Publishing, Copenhagen).

Sakakibara, K.-I. (**2003**). "Production mechanism of voice quality in singing," Journal of the Phonetic Society of Japan **17**, 27-39.

Sakakibara, K.-I., Fuks, L., Imagawa, H., and Tayama, N. (**2004**). "Growl Voice in Ethic and Pop Styles," in *Proceedings of the International Symposium on Musical Acoustics (ISMA2004)* (Nara, Japan).

Salomão, G. L. (**2008**). "Relation between perceived voice register and flow glottogram

parameters in males " J Acoust Soc Am **124**, 546-551.

Sapienza, C., Stathopoulos, E. T., and Dromey, C. (**1998**). "Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: effects of measurement criteria and sound pressure level," J Voice **12**, 31-43.

Sataloff, R. T. (**1997**). "Clinical anatomy and physiology of the voice," in *Professional voice: the science and art of clinical care*, edited by R. T. Sataloff (Singular Publishing Group, San Diego, CA), pp. 111-130.

Sataloff, R. T. (**1998**). *Vocal health and pedagogy* (Singular Publishing Group, San Diego, California).

Scherer, R. C., Druker, D. G., and Titze, I. R. (**1988**). "Electroglottography and direct measurement of vocal fold contact area," in *Vocal Fold Physiology, Vol.2: Voice production, mechanisms and functions*, edited by O. Fujimura (Ed. Raven Press, New York), pp. 279-290.

Schutte, H. K., and Miller, D. G. (**2001**). "Measurement of closed quotient in a female singing voice by electroglottography and videokymography," in *Vth International Conference Advances in Quantitative Laryngology, Groningen, the Netherlands, April 27-28, 2001*, edited by H. K. Schutte (Groningen Voice Research Lab, University of Groningen).

Seidner, W., and Wendler, J. (**2004**). *Die Sängerstimme* (Henschel Verlag, Berlin).

Sissom, D., Rice, D., and Peters, G. (**1991**). "How cats purr," The Zoological Society of London **223**, 67-78.

Smith, A., and Childers, D. (**1983**). "Laryngeal evaluation using features from speech and the electroglottograph," IEEE Trans Biomed Eng **30**, 755-759.

Smith, J. O. I. (**2007**). *Mathematics of the Discrete Fourier Transform (DFT), with Audio Applications* (W3K Publishing).

Södersten, M., Hertegard, S., and Hammarberg, B. (**1995**). "Glottal Closure, Transglottal Airflow, and Voice Quality in Healthy Middle-Aged Women," J Voice **9**, 182-197.

Sondhi, M. (**1975**). "Measurement of the glottal waveform," J Acoust Soc Am **57**, 228-232.

Sonesson, B. (**1960**). "On the anatomy and vibratory pattern of the human vocal folds. With special reference to a photo-electrical method for studying the vibratory movements," Acta Oto-Laryngologica (Stockholm) **(suppl 156)**.

Stark, J. (**1999**). "Registers: Some Tough Breaks," in *Bel Canto - A History of Vocal Pedagogy* (University of Toronto Press Incorporated), pp. 57-90.

Story, B. (**2002**). "An overview of the physiology, physics and modeling of the sound source for vowels," Acoust Sci Tech **23**.

Story, B. (**2004**). "Vowel acoustics for speaking and singing," Acta Acostica united with Acostica **90**, 629-640.

Sundberg, J. (**1972**). "An articulatory interpretation of the singing formant," STL-QPSR, Stockholm **1**, 45-53.

Sundberg, J. (**1974**). "Articulatory interpretation of the "singing formant"," J Acoust Soc Am **55**, 838-844.

Sundberg, J. (**1977**). "The acoustics of the singing voice," Scientific American **236**, 82-91.

Sundberg, J. (**1981a**). *Research Aspects on Singing* (Royal Swedish Academy of Music, Stockholm).

Sundberg, J. (**1981b**). "The voice as a sound generator," in *Research aspects on singing* (Royal Swedish Academy of Music, Stockholm, Sweden), pp. 6-14.

Sundberg, J., and Skoog, J. (**1997**). "Dependence of jaw opening on pitch and vowel in singers," J Voice **11**, 301-306.

Švec, J. G., and Granqvist, S. (**2010**). "Guidelines for Selecting Microphones for Human Voice Production Research," American Journal of Speech-Language Pathology **19**, 356-368.

Švec, J. G., and Schutte, H. (**1996**). "Videokymography: High-speed line scanning of vocal fold vibration," J Voice **10**, 201-205.

Švec, J. G., Horacek, J., Šram, F., and Vesely, J. (**2000**). "Resonance properties of the vocal folds: in vivo laryngoscopic investigation of the externally excited laryngeal vibrations," J Acoust Soc Am **108**, 1397-1407.

Švec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: data from living and excised human larynges," J Acoust Soc Am **106**, 1523-1531.

Švec, J. G., Šram, F., and Schutte, H. K. (**2007**). "Videokymography in Voice Disorders: What to Look For?" Ann Otol Rhin Laryngol**116**, 172-180.

Švec, J. G., Šram, F., and Schutte, H. K. (**2009**). "Videokymography," in *The Larynx*, edited by M. Fried, and A. Ferlito (Plural Publishing, San Diego, CA).

Švec, J. G., Šram, F., and Schutte, H. K. (**2001**). "Development and Application of Videokymography for High-Speed

Examination of Vocal-Fold Vibration," in *LP'2000*, edited by B. Palek, and O. Fujimura (Prague, Czech Republic).

Švec, J. G., Sundberg, J., and Hertegard, S. (**2008**). "Three registers in an untrained female singer analyzed byvideokymography, strobolaryngoscopy and sound spectrography," J Acoust Soc Am **123**, 347-353.

Teaney, D., Fourcin, A. (**1980**). "The electrolaryngograph as a clinical tool for the observation and analysis of vocal fold vibration," in *Ninth Symposium Care of the Professional Voice* (The Voice Foundation, The Juilliard School, New York City), pp. 128 - 134.

Titze, I. R. (**1983**). "The physics of small-amplitude oscillation of the vocal folds," J Acoust Soc Am **4**, 1536-1552.

Titze, I. R. (**1988a**). "A Framework for the Study of Vocal Registers," J Voice **2**, 183-194.

Titze, I. R. (**1988b**). "The physics of small-amplitude oscillation of the vocal folds," J Acoust Soc Am **83**, 1536-1552.

Titze, I. R. (**1989**). "A Four-parameter Model of the Glottis and Vocal Fold Contact Area," Speech Comm **8**, 191-201.

Titze, I. R. (**1990**). "Interpretation of the electroglottographic signal," J Voice **4**, 1-9.

Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech).

Titze, I. R. (**2004a**). "A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice," J Voice **18**, 292-298.

Titze, I. R. (**2004b**). "Theory of Glottal Airflow and Source-Filter Interaction in Speaking and Singing," Acta Acostica united with Acostica **90**, 641-648.

Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: theory," J Acoust Soc Am **123**, 2733-2749.

Tsai, C., Yio-Wha, S., Hon-Man, L., and Tzu-Yu, H. (**2006**). "Laryngeal Mechanisms During Human 4-kHz Vocalization Studied With CT, Videostroboscopy, and Color Doppler Imaging " JVoice **22**, 275-282.

Unger, J., Meyer, T., Herbst, C. T., Döllinger, M., and Lohscheller, J. (**2011**). "PVG-Wavegrams: Three-dimensional visualization of vocal fold dynamics," in *7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)* (Firenze, IT).

van den Berg, J. (**1958**). "Myoelastic-aerodynamic theory of voice production," J Speech Hearing Research **3**, 227-244.

van den Berg, J., and Tan, T. S. (**1959**). "Results of experiments with human larynxes," Practica Oto-Rhino-Laryngologica **21**, 425-450.

van den Berg, J., Vennard, W., Burger, D., and Shervanian, C. C. (**1960**). *Voice production. The vibrating larynx. (Instructional film)* (University of Groningen).

van den Berg, J. (**1963**). "Vocal ligaments versus registers," The NATS Bulletin, 16-31.

Verikas, A., Uloza, V. D., Bacauskiene, M., Gelzinis, A., and Kelertas, E. (**2009**). "Advances in laryngeal imaging," Eur Arch Otorhinolaryngol **266**, 1509-1520.

Vilkman, E., Alku, P., and Laukkanen, A. M. (**1995**). "Vocal-fold collision mass as a differentiator between registers in the low-pitch range," J Voice **9**, 66-73.

Wendler, J. (**2008**). "Singing and science," Folia Phon **60**, 279-287.

Zemlin, W. (**1998**). "Phonation," in *Speech and Hearing Science - Anatomy and Physiology*, edited by W. Zemlin (Allyn and Bacon, Boston), pp. 100-196.

Ziethe, A., Patel, R., Kunduk, M., Eysholdt, U., and Graf, S. (**2011**). "Clinical analysis methods of voice disorders," Current Bioinformatics **6**, 270-285.

Supplement A: Manuscript I

Electroglottographic wavegrams: a technique for visualizing vocal fold dynamics noninvasively

C. T. HERBST, W. T. FITCH, J. G. ŠVEC

Reprinted with permission from
J Acoust Soc Am **128**, 3070-3078 (**2010**).
 [DOI: 10.1121/1.3493423]

Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively[a)]

Christian T. Herbst[b)]
Laboratory of Biophysics, Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, tř. 17. listopadu 12, 771 46 Olomouc, Czech Republic

W. Tecumseh S. Fitch
Laboratory of Bio-Acoustics, Department of Cognitive Biology, University of Vienna, Althanstrasse 14, 1090 Wien, Austria

Jan G. Švec
Laboratory of Biophysics, Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, tř. 17. listopadu 12, 771 46 Olomouc, Czech Republic

(Received 4 June 2010; revised 5 August 2010; accepted 25 August 2010)

A method for analyzing and displaying electroglottographic (EGG) signals (and their first derivative, DEGG) is introduced: the electroglottographic wavegram ("wavegram" hereafter). To construct a wavegram, the time-varying fundamental frequency is measured and consecutive individual glottal cycles are identified. Each cycle is locally normalized in duration and amplitude, the signal values are encoded by color intensity and the cycles are concatenated to display the entire voice sample in a single image, similar as in sound spectrography. The wavegram provides an intuitive means for quickly assessing vocal fold contact phenomena and their variation over time. Variations in vocal fold contact appear here as a sequence of events rather than single phenomena, taking place over a certain period of time, and changing with pitch, loudness and register. Multiple DEGG peaks are revealed in wavegrams to behave systematically, indicating subtle changes of vocal fold oscillatory regime. As such, EGG wavegrams promise to reveal more information on vocal fold contacting and de-contacting events than previous methods.
 [DOI: 10.1121/1.3493423]

PACS number(s): 43.70.Jt, 43.70.Gr, 43.75.Rs [AL] Pages: 3070–3078

I. INTRODUCTION

Electroglottography was invented by Fabre (1957) to monitor the vibration of vocal folds *in vivo*. A low intensity, high-frequency current is passed between two electrodes that are placed externally on the neck, on each side of the thyroid cartilage at vocal fold level. The contacting and de-contacting of the vocal folds causes variations in the electrical impedance across the larynx, resulting in a variation in current between the two electrodes (Baken, 1992; Fourcin and Abberton, 1971). This variation in the current flow has been found to be related to changes in vocal fold contact area (Scherer *et al.*, 1988).

The electroglottographic signal is a time-varying one-dimensional representation of the complex three-dimensional motion of the vocal folds. Experimental research (Baer *et al.*, 1983; Childers *et al.*, 1983) has confirmed a close relation between peaks in the derivative of the EGG signal and the contacting and de-contacting events of the vocal folds. It has been shown that landmarks in the EGG waveform are related to the movement and position of the vocal folds during phonation (Hess and Ludwigs, 2000; Rothenberg, 1979). The physiologic relevance of the EGG signal has been examined theoretically by Titze (1989, 1990), who discussed the effects of (a) increased glottal adduction; (b) glottal convergence and vertical phasing; (c) medial vocal fold surface bulging; and (d) increased vertical phasing in vocal fold vibration.

The first mathematical derivative of the EGG waveform (DEGG) reflects the rate of change of the EGG with time (Childers and Krishnamurthy, 1985; Teaney and Fourcin, 1980). The timing of glottal closure can often be derived from a single maximum in the DEGG signal (Childers *et al.*, 1983). However, in some subjects, multiple peaks for both the contacting and the de-contacting phase can be seen. It has been suggested that the multiple peaks in the DEGG signal might be related to zipper-like opening and closing of the vocal folds (either anterior-to-posterior or posterior-to-anterior) (Henrich *et al.*, 2004; Hess and Ludwigs, 2000). It is also conceivable that they result from artifacts, such as mucus strands (Colton and Conture, 1990), or abnormalities in vocal fold tissue structure distorting the regularity of the EGG waveform (Baken and Orlikoff, 2000; Kitzing, 1990). Such DEGG multiple peaks phenomena have, however, not yet been investigated in detail.

Quantitative analysis of the EGG waveform has been achieved by measuring the relative proportion of glottal

[a)]Portions of this work were presented at the Fifth International Conference on the Physiology and Acoustics of Singing (PAS5), 2010, Stockholm, Sweden.
[b)]Author to whom correspondence should be addressed. Electronic mail: herbst@ccrma.stanford.edu

closure within a glottal vibratory period (Rothenberg and Mahshie, 1988), known as the "larynx closed quotient" (Howard, 1995) or "contact quotient" (Orlikoff, 1991) (CQ_{EGG}). This quotient has been found useful in clinical as well as basic voice research (Schutte and Miller, 2001; Švec *et al.*, 2008). A sudden decrease of CQ_{EGG} has been reported, e.g., during the transition from chest/modal to falsetto register in singing (Henrich *et al.*, 2005; Miller *et al.*, 2002). Research has shown, however, that the CQ_{EGG} is dependent on the choice of algorithm used to determine the contacting and de-contacting events, and must therefore be used with caution (Herbst and Ternström, 2006; Higgins and Schulte, 2002; Kania *et al.*, 2004; Sapienza *et al.*, 1998).

In order to study longer sequences of speech and singing, hundreds of glottal cycles need to be examined, especially when investigating phonations with changing physiologic parameters. In such cases, entire EGG signals are often represented by time-varying analysis parameters (e.g., EGG contact quotient or EGG signal amplitude). Since a single analysis parameter does not reflect all nuances seen in the EGG waveform, valuable information might be disregarded in the signal processing stage.

The need of extracting meaningful information from large amounts of data has been addressed by several authors. For better visualization of signals conveying information concerning phonation, methods have been conceived which segment the signal into single glottal cycles and show how these cycles change over time. Among those techniques are: the "dreidimensionale Periodizitätsanalyse" (3D-PAN), for analyzing acoustic signals (Sedlacek and Michek, 1988); the voice cascadation method for analysis of acoustic and EGG signals (Berg and Gall, 1999a, 1999b); videokymography (VKG) for documenting vibratory pattern of the vocal folds (Švec and Schutte, 1996; Švec *et al.*, 2007); and phonovibrography for reducing the amount of data gathered in high-speed imaging (Lohscheller *et al.*, 2008). The advantage of these approaches is that ongoing developments and gradual changes of laryngeal behavior can be seen in one single image.

In the present study, a similar approach for data reduction of the electroglottographic signal will be presented. A new method for analyzing and displaying EGG signals and their first derivative (DEGG) is introduced, which (a) allows monitoring the EGG (or DEGG) signal over time within a single image; and (b) provides an intuitive means for quickly assessing the vocal fold contact phase and its variation over time.

II. ASSEMBLING THE WAVEGRAM: METHOD

An algorithm has been developed to show how the EGG waveform changes over time, from cycle to cycle, in one single image (called a "wavegram"). An EGG signal (or a portion thereof) is taken and converted into a graph where time is displayed on the *x*-axis from left to right; the progress of consecutive glottal cycles is displayed on the *y*-axis from bottom to top; and the locally normalized vocal contact area is encoded in the z-axis as a color intensity value. The algorithm (implemented with a C++ signal processing library written by author CH) consists of the following stages: (1) detection and separation of EGG cycles, (2) normalization of the amplitudes of the EGG cycles, (3) coding the normalized EGG values with color intensity, (4) concatenation of the color-intensity-coded EGG cycles, and (5) normalization of the period length to obtain a uniform graph display. These steps are described below in detail.

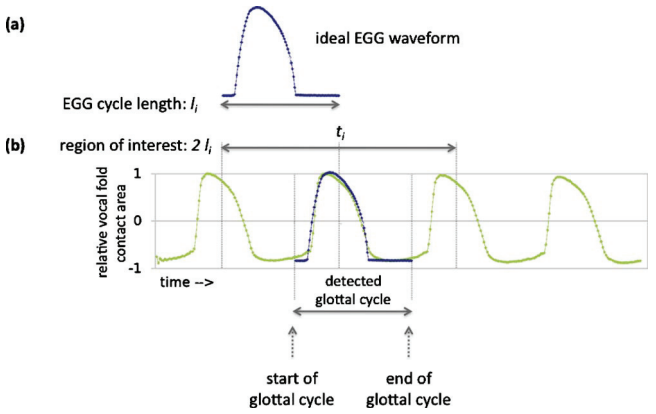


FIG. 1. (Color online) Extraction of one glottal cycle. Ideal EGG waveform templates are (a) stretched to the estimated glottal cycle duration, and (b) correlated with a portion of the EGG signal having twice the size of the estimated glottal cycle (the "region of interest"). The beginning of the glottal cycle is defined as the beginning of the ideal waveform fit.

A. Detection of glottal cycles

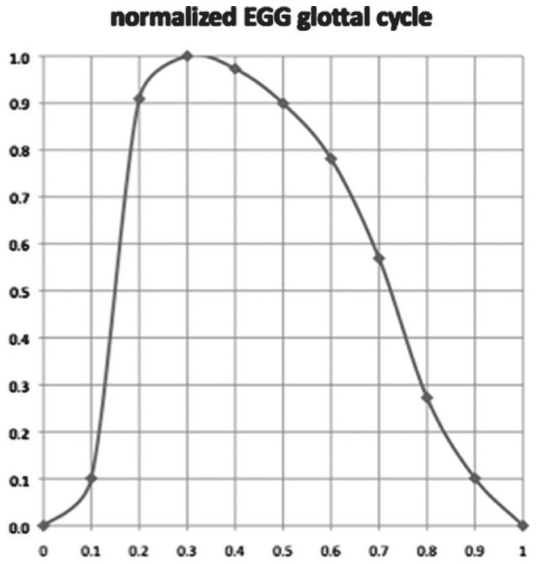
To detect glottal cycles in quasi-periodic phonation, the period duration of the input signal is initially determined by performing an auto-correlation analysis as described by Boersma (1993). The part of the EGG (or DEGG) signal that represents single glottal cycles is defined and determined by repeatedly cross-correlating an ideal EGG waveform "template" [taken from Titze (1990)] with the actual EGG (or DEGG) signal (see Fig. 1). During this process, the templates are first stretched to the approximate cycle duration (as determined by the initial auto-correlation analysis), and then each template is correlated with a region of the EGG signal having twice the size of the computed glottal cycle duration. The actual beginning of each glottal cycle is defined as the beginning of the ideal fit.

For non-periodic phonation and EGG waveforms deviating from the ideal EGG waveform template (e.g., samples obtained by dysphonic speakers), the method as described above might fail. In such a case, the period should be rather determined on a cycle-to-cycle basis from direct inspection of the electroglottographic signal (and its first derivative) in the time domain (Fourcin and Abberton, 2008). Such an alternative algorithm for glottal cycle detection is described in the Appendix.

B. Normalization and color-coding

Every single extracted glottal cycle is locally normalized in amplitude (ordinate). The amplitude values (y_j) are then coded into monochrome color information:

$$col_j = 255(1 - y_j), \quad j \in [1 \,..\, n], \tag{1}$$



data point	0	1	2	3	4	5	6	7	8	9	10
xRel	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
yRel	0.00	0.10	0.91	1.00	0.97	0.90	0.78	0.57	0.27	0.10	0.00
color value	0	26	232	255	248	229	199	145	70	26	0

color

FIG. 2. Conversion of normalized EGG signal data (y-axis values) into color information in order to create the temporary graphs (see text for details). In this simplified example only 11 samples are considered, while typically 40–400 samples would be used (depending on fundamental frequency and sampling rate).

where n is the size of the glottal cycle, $0 \leq y_j \leq 1$ is the locally normalized EGG amplitude (which is proportional to the relative vocal fold contact area); and col_j is the resulting monochrome color information. Here, high values correspond to dark colors. Apart from the normalization, the color coding process is, in principle, similar to the one used for converting the consecutive sound spectra to sound spectrograms (Koenig *et al.*, 1946). The process and the resulting color-coded EGG cycle is displayed in Fig. 2.

C. Cycle-concatenation and normalization of the final display

The color-coded strips, corresponding to individual glottal cycles are rotated by 90° counter-clockwise [Fig. 3(c)]. The height of the individual cycle plots corresponds now to the period duration, and a specific point along the y-axis represents a particular phase of the glottal cycle. The width of the individual strips can be proportionally squeezed to display as many waveforms as possible, allowing us to represent the whole phonation in a single image. As a final step, the heights of the individual cycle plots (i.e., period durations) are normalized by means of interpolation to form the final graph. As shown later, this height normalization is useful for investigating modifications of the events occurring within the glottal cycle. In the resulting wavegram, time is displayed on the x-axis; normalized progress of consecutive glottal cycles is displayed on the y-axis; and normalized vo-

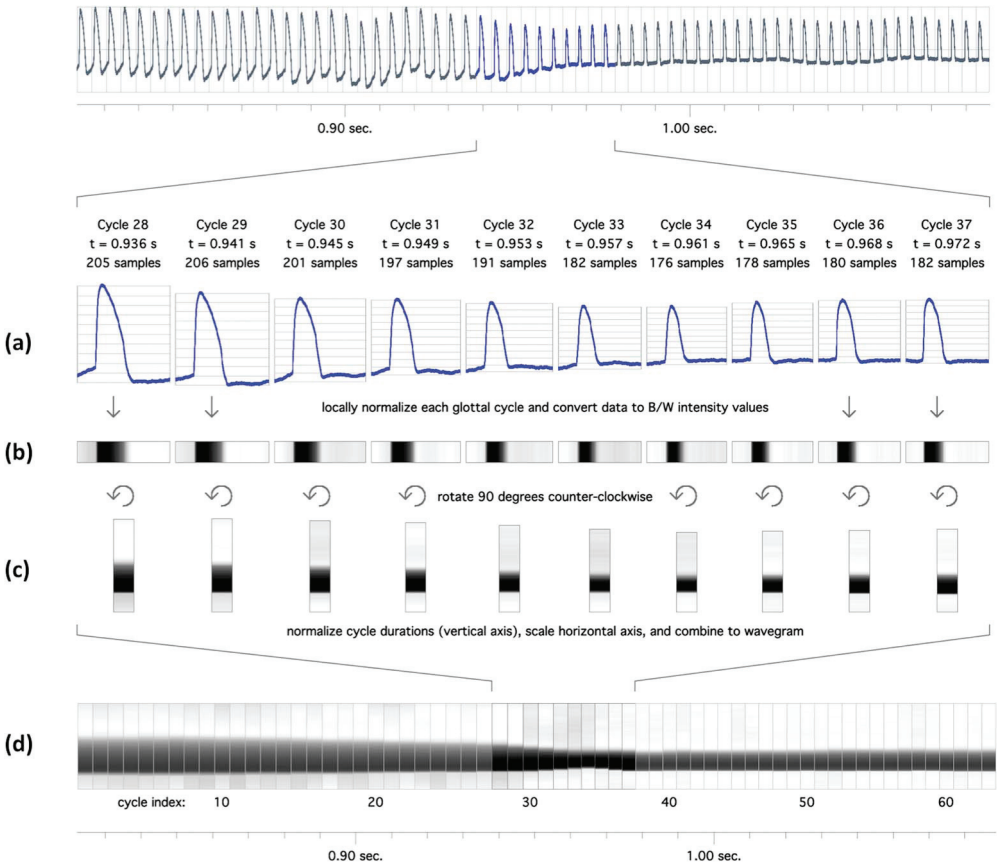


FIG. 3. (Color online) Illustration of the basic processing steps to create an electroglottographic wavegram: consecutive glottal cycles of an EGG signal are extracted (a). The locally normalized data values are converted into monochrome color information, and are plotted as a strip representing one glottal cycle each (b). All strips are rotated 90° counter-clockwise (c). Glottal cycle duration is normalized by scaling the individual glottal cycle plots to the same height, and the resulting graphs are combined to form the final display, the EGG wavegram (d), and the resulting graphs are combined to form the final display, the EGG wavegram (e).

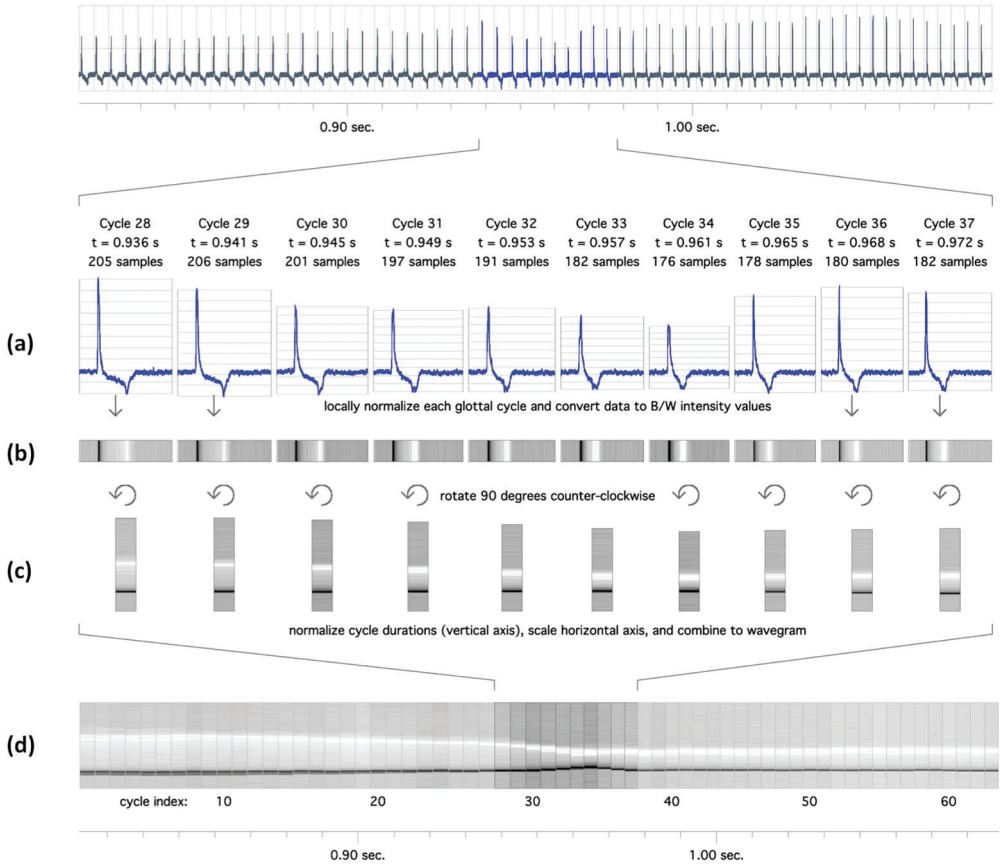


FIG. 4. (Color online) As Fig. 3, but with a DEGG signal as input.

cal fold contact area within a cycle is shown on the z-axis by means of (monochrome) color intensity [see Fig. 3(d)].

As an alternative display option, the first derivative of the EGG signal (DEGG) can be used as an input (rather than the EGG signal itself) to the algorithm of generating the wavegram—see Fig. 4. DEGG-based wavegrams provide a clearer view of the moments of most rapid change in the vocal fold contact area and allow us to track multiple peaks in both the contacting and de-contacting phase of the DEGG-signal (see later in Fig. 7).

In order to reduce the resulting graph width, an optional data reduction task can be performed by only considering glottal cycles at user-defined time intervals. This method of data reduction was used for the examples shown in Figs. 6–10 in Sec. III and the Appendix of this manuscript.

III. RESULTS

Wavegrams for a sustained stable phonation (sung by a male singer) are shown in Fig. 5 on the right. On the left side of the figure, there are the corresponding EGG (dark color) and DEGG (light color) waveforms. The following landmarks (Childers and Krishnamurthy, 1985; Rothenberg, 1979) for the typical EGG waveform are identified in the figure:

(1) Start of the glottal cycle (x=0); this landmark has been chosen arbitrarily around the moment of minimum vocal fold contact area in order to be able to display the entire contact phase of the glottal cycle in the EGG wavegram.
(2) Initiation of glottal closure (x=0.13), identified by a minor, but pronounced change in EGG signal steepness and a secondary positive peak in the DEGG signal. The EGG wavegram shows a minute change in color intensity (light to dark in upward vertical dimension). In the DEGG wavegram this phenomenon is documented as a fine line of gray color.
(3) Maximum increase of vocal fold contact (x=0.18), identified by the positive peak of the DEGG signal. The EGG wavegram exhibits an abrupt change in color intensity (light to dark in upward vertical dimension). In the DEGG wavegram the positive peak of the DEGG signal appears as a line of dark gray color.
(4) Maximum decrease of vocal fold contact (x=0.57), identified by the negative peak of the DEGG signal. This event in the glottal cycle appears in the DEGG wavegram as a bright line. In the EGG wavegram, this event is not clearly visible, since the de-contacting phase of a glottal cycle is distributed over a longer period of time.
(5) End of glottal cycle.

The temporal locations of the strongest positive and negative peaks of the DEGG signal can be used to calculate the EGG contact quotient (Henrich *et al.*, 2004). Other methods of calculating the EGG contact quotient, e.g., based on a criterion level, can also be used. One has to be aware, however, that these different methods are likely to produce different results, based on algorithm settings (Herbst and Ternström, 2006). When applying the DEGG criterion to the

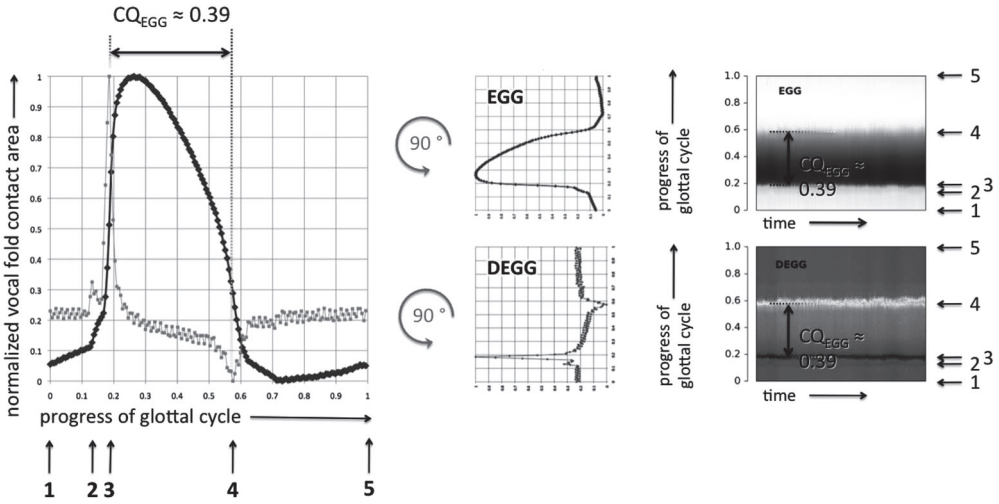


FIG. 5. Typical landmarks in the EGG wavegram, both EGG-based (top right) and DEGG-based (bottom right), as related to the EGG signal of a single glottal cycle and its first derivative (left and center): (1) beginning of glottal cycle, arbitrarily chosen in order to display the entire contact phase of the glottal cycle; (2) initiation of vocal fold closure; (3) maximum rate of increase of vocal fold contact; (4) maximum rate of decrease of vocal fold contact; (5) end of glottal cycle. See text for further details. Note that the contact quotient (CQ_{EGG}) was calculated based on the strongest positive and negative peak in the first derivative of the EGG waveform. Other methods of calculating the EGG contact quotient, e.g., based on a criterion level, are likely to produce different results, based on individual algorithm settings.

glottal cycle displayed in Fig. 5, a contact quotient of ca. 0.39 is calculated. The DEGG wavegram shows this clearly, since the positive and negative peak of the DEGG signal are displayed as distinct lines of darker and lighter color, respectively.

In Fig. 6, phonation sustained at a stable fundamental frequency of approximately 233 Hz with increasing vocal intensity is shown (female classical singer, 4 years of academic vocal training). The DEGG wavegram [Fig. 6(c)] reveals a steady change from phonation with a shorter duration of glottal closure to phonation with a longer duration of glottal closure (see Fig. 6(a): EGG waveforms extracted at $t=1.5$ s and $t=5$ s, respectively). The gradual change in both the contact phase and the amplitude of the EGG waveform [Fig. 6(e)] suggests that no abrupt change in vocal fold vibratory pattern has occurred (Roubeau *et al.*, 2009).

Phonation involving increasing and decreasing vocal intensity (a so-called messa-di-voce) produced at a stable pitch (F#3, ca. 185 Hz) by an untrained male amateur singer is shown in Fig. 7. The most significant difference from the phonation displayed in Fig. 6 is the presence of multiple converging DEGG peaks in the contacting phase of vocal fold vibration at lower intensity levels [Fig. 7(c)]. At $t \approx 4.5$ s, after the converging DEGG peaks have fully merged, a sudden increase of vocal fold contact phase occurred. This is reflected by the fact that the strongest negative peak of the DEGG signal [Fig. 7(a)] extracted at $t=5$ s occurred at a later stage of the glottal cycle, as compared to the strongest negative DEGG peak of a glottal cycle extracted at $t=4.3$ s. A reversed phenomenon was seen around $t \approx 7.5$ s, where an abrupt shortening of glottal contact phase is accompanied by the separation of multiple DEGG peaks in the contacting phase (see ellipses in Fig. 7). This relationship between multiple peaks and other aspects of vocal fold behavior, as often observed in wavegrams, suggests that such peaks reveal subtle aspects of vocal fold dynamics.

In Fig. 8, phonation with increasing fundamental frequency (ca. 208–415 Hz) is illustrated in a wavegram (female classical singer, 2 years of academic vocal training). Analysis data revealed a sudden change of EGG contact quotient around $t \approx 0.95$ s. This coincided with (i) a sudden

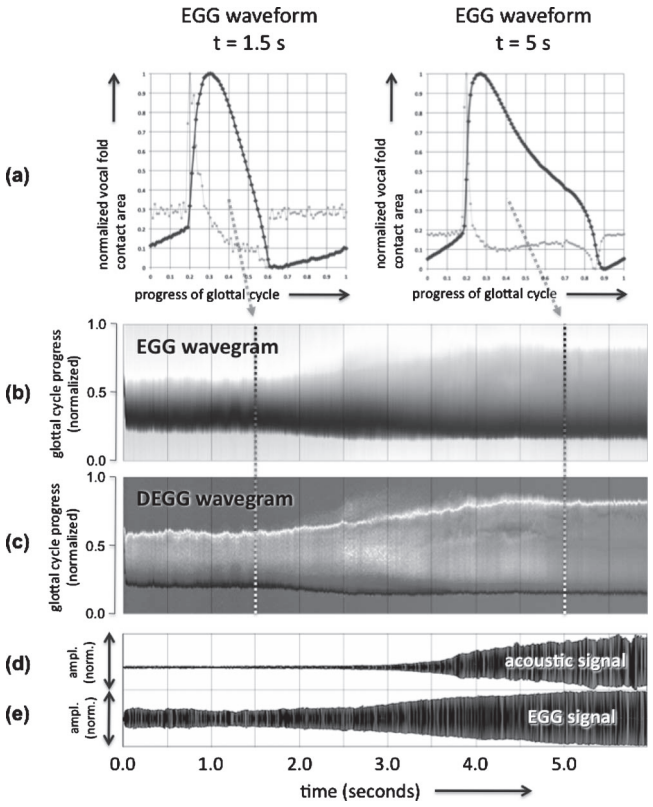


FIG. 6. Changes of the EGG and audio signals in case of female phonation with increasing vocal intensity. (a) EGG (black) and DEGG (gray) waveforms representing glottal cycles extracted at $t=1.5$ s and $t=5$ s, respectively; (b) EGG wavegram; (c) DEGG wavegram; [(d) and (e)] amplitude plot of audio and EGG signal. The wavegrams [(b) and (c)] reveals a gradual increase of vocal fold contact phase.

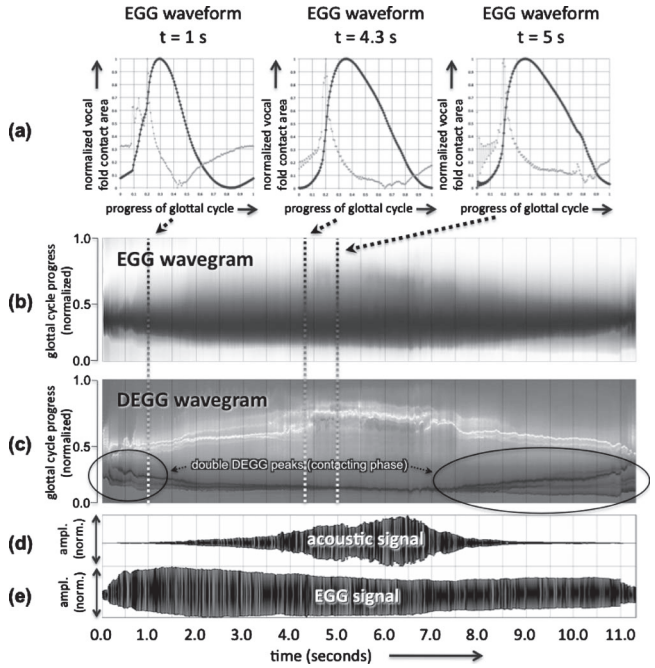


FIG. 7. Changes of the EGG and audio signals in a case of male phonation with increasing and decreasing vocal intensity. (a) EGG (black) and DEGG (gray) waveforms representing glottal cycles extracted at $t=1$ s, $t=4.3$ s and $t=5$ s, respectively; (b) EGG wavegram; (c) DEGG wavegram; [(d) and (e)] amplitude plot of audio and EGG signal. Multiple converging/diverging DEGG peaks are seen at lower intensity levels (see text for details).

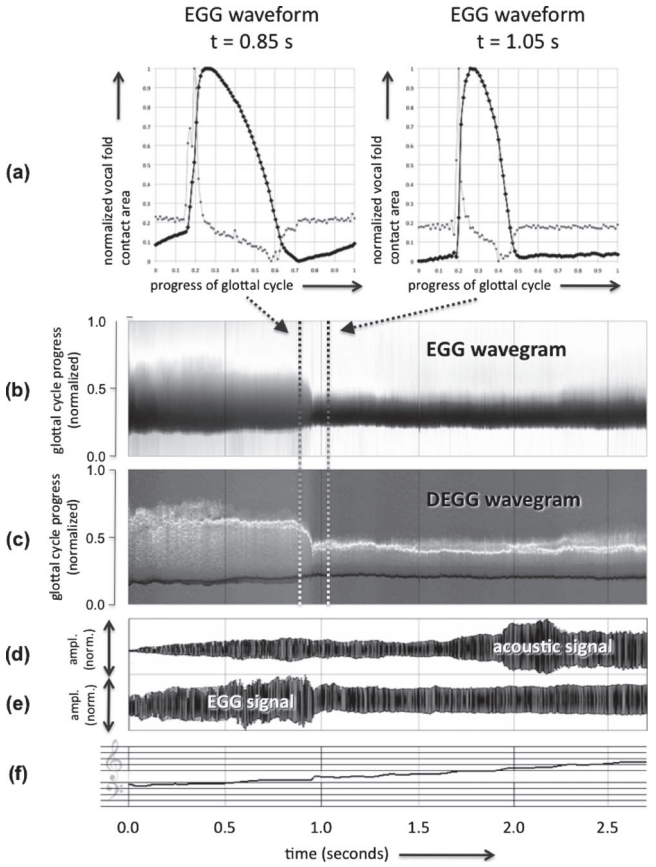


FIG. 8. Changes of the EGG and audio signals in case of female phonation with increasing fundamental frequency. (a) EGG (black) and DEGG (gray) waveforms representing glottal cycles extracted at $t=0.85$ s and $t=1.05$ s, respectively; (b) EGG wavegram; (c) DEGG wavegram; [(d) and (e)] amplitude plot of audio and EGG signal; (f) fundamental frequency displayed in musical notation (ca. 208–415 Hz). The data reveals an abrupt transition from chest to falsetto register at $t\approx0.95$ s, indicated by (i) a sudden decrease of glottal contact duration (wavegram); (ii) a sudden decrease of EGG signal amplitude; and (iii) an abrupt pitch shift.

change of vocal fold contact phase [EGG waveforms and wavegrams, Figs. 8(a)–8(c)]; (ii) a sudden change of overall EGG signal amplitude [Fig. 8(e)] (Roubeau *et al.*, 2009); (iii) an abrupt pitch change [Fig. 8(f)], that was accompanied by an audible change of vocal timbre; and (iv) a sudden decrease of sound pressure level by 6 dB (not shown in figure). The change of EGG waveform happened over a period of ca. 0.028 s (7 glottal cycles).

Phonation with increasing fundamental frequency (ca. 208–415 Hz) with a continuous register transition is shown in Fig. 9 (female classical singer, 4 years of academic vocal training). The wavegram [Figs. 9(b) and 9(c)] exhibited no abrupt change of glottal contact phase. This was corroborated by a stable overall EGG signal amplitude [Fig. 9(e)] and smooth changes of fundamental frequency [Fig. 9(f)]. Careful examination of the acoustic signal revealed a barely audible timbral change around t=2 s, which did, however, not coincide with a significant change of sound pressure level. The EGG waveform [Fig. 9(a)] gradually changed from a chest phonation (t=0.5 s) with a strong "knee," i.e., a "bump" in the de-contacting phase of the EGG waveform where the signal amplitude is beginning to decrease more rapidly (Hess and Ludwigs, 2000), to a falsetto-like waveform (t=3.5 s, see Fig. 9). In particular, the knee was gradually dissipating. From $t\approx0.8$ s to $t\approx2.8$ s, two negative DEGG peaks were seen. The evidence suggests that—unlike the phonation shown in Fig. 8—no abrupt change of vocal fold oscillatory regime took place, i.e., the transition from chest to falsetto register was accomplished gradually, as intended.

IV. DISCUSSION AND CONCLUSIONS

The wavegram technique provides a new and potentially powerful method for displaying entire electroglottographic signals, or parts thereof. Much like in the spectrogram, information on vibratory behavior developing in time is compacted into one single graph providing insight into changes of vocal fold dynamics. Nevertheless, waveform details of individual glottal cycles (and their gradual development over time) are preserved, thus providing a useful tool to quickly gain physiologic insights into longer bouts of phonation than are visible in a simple amplitude graph.

The wavegram reveals changes of vocal fold contact phase, as well as of phenomena that cannot easily be seen with other methods for displaying electroglottographic signals/waveforms. Those phenomena, which develop over time, are expected to be related to physiologic behavior of vocal fold vibration, the nature of which has not yet been fully explored.

When analyzing electroglottographic signals, derivative analysis parameters such as the contact (CQ) or the contact index (CI) are calculated. Those consist of only one time-varying variable, the calculation of which is dependent on

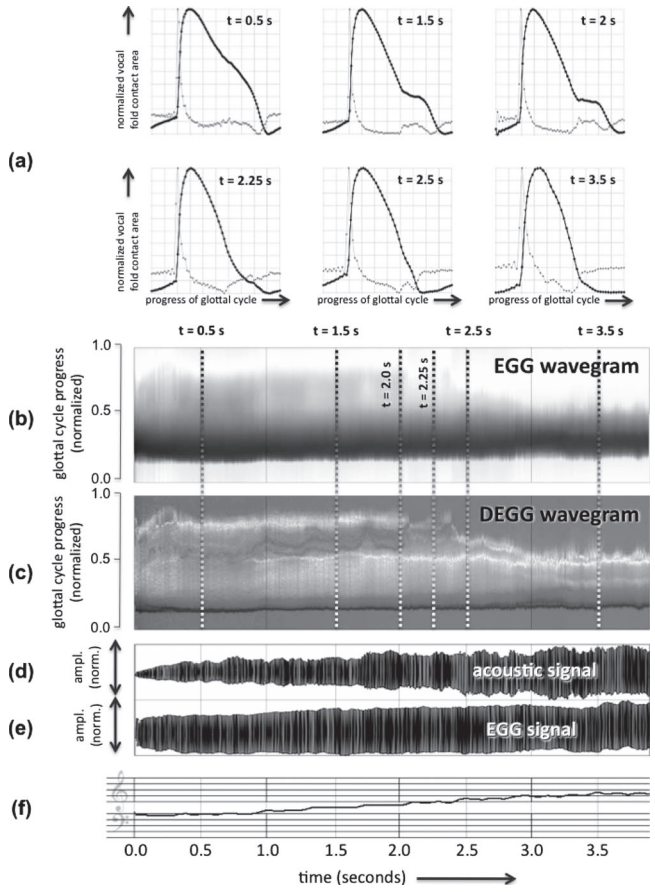


FIG. 9. Changes of the EGG and audio signals in case of female phonation with increasing fundamental frequency. (a) EGG (black) and DEGG (gray) waveforms representing glottal cycles extracted at t=0.5 s, t=1.5 s, t=2 s, t=2.25 s, t=2.5 s, and t=3.5 s, respectively; (b) EGG wavegram graph; (c) DEGG wavegram graph; [(d) and (e)] amplitude plot of audio and EGG signal; (f) fundamental frequency displayed in musical notation (ca. 208–415 Hz). The data reveals a smooth transition from chest to falsetto register, indicated by (i) a gradual change of the time-varying EGG waveform representing one glottal cycle; (ii) a lack of abrupt change of glottal closure duration in the wavegram; (iii) a steady EGG signal amplitude; and (iv) no abrupt (involuntary) pitch changes.

user inputs which are partially arbitrary, e.g., the specification of a certain threshold value (Herbst and Ternström, 2006). In the wavegram, on the other hand, the EGG waveform is treated "as is." Apart from period and alignment of glottal cycles, no absolute values are calculated. Thus, the wavegram constitutes an alternative to the contact quotient monitoring, since it provides additional information, and does not depend on any arbitrary threshold criterion. Our observations suggest that this can provide novel insights into details of vocal fold behavior that are easily overlooked in other analysis techniques.

As a prominent example, converging multiple DEGG peaks in the contacting phase appear to constitute a systematic and consistent phenomenon seen in a considerable proportion of subjects (an example is shown in Fig. 7). The same is true for multiple DEGG peaks in the de-contacting phase, which may be related to gradual changes in the vocal fold oscillatory regime. The consistent behavior of these peaks over time provides evidence that they reveal physiologic phenomena, and should not be considered artifacts in the strict sense. In particular, multiple DEGG peaks are induced by vocal fold vibratory phenomena during which vocal fold contact abruptly increases (in the contacting phase) or decreases (in the de-contacting phase). The exact universal interpretation of these phenomena has not yet been available. It can be speculated that they are related to phase differences (in the superior-inferior as well as the anterior-posterior vocal folds dimension), commonly termed "zipper-like" vocal fold opening or closure (Childers *et al.*, 1986). Further research investigating the physiological relevance of these phenomena is warranted.

The wavegram reveals that vocal fold contacting and de-contacting "events" are more complex than commonly assumed. Rather than a single incident, vocal fold contacting and de-contacting should be considered a sequence of events, taking place over a certain period of time. The analyses presented here indicate that this sequence of events can change with fundamental frequency, loudness and register. The EGG signal appears to convey more physiological information on vocal fold contacting and de-contacting events than what is offered by more traditional representations of the EGG signal. The wavegram technique promises to provide a method to visualize, further explore and understand this "hidden" information.

The wavegram method focuses on the contacting and de-contacting phenomena occurring within individual glottal cycles and monitors their changes as the phonation progresses. As such, the method is not primarily intended for detecting pathologies, but rather for better understanding the physiological phenomena occurring during vocal fold vibration. The potential limitations of the wavegram technique in its current state are twofold: (a) It is dependent on proper period detection, which is problematic in non-periodic/pathologic phonations (Titze, 1995); (b) The wavegram is primarily intended for visual analysis, and it offers no automatic quantitative measurement for relevant signal fluctuations. A future version of the algorithm might offer an option to skip the normalization of the glottal period duration, thus revealing period perturbations such as jitter (Vieira *et al.*, 1996; Vieira *et al.*, 2002), shimmer or vibrato. Further research including experiments with recordings from dysphonic speakers is necessary in order to test the applicability of the wavegram technique in pathologic voices.

Electroglottography is a relatively inexpensive and non-invasive technique. It holds a great potential for singing pedagogy (Herbst *et al.*, 2010; Howard *et al.*, 2004; Miller and Schutte, 1999; Rossiter *et al.*, 1996), and possibly for detecting voice disorders (Baken and Orlikoff, 2000; Smith and Childers, 1983) and in voice therapy (Baken and Orlikoff, 2000; Fourcin *et al.*, 1995). The novel wavegram technique, presented in this paper, enables us to better tap the rich potential of the EGG waveform. Wavegrams allow assessment and analysis of the EGG waveforms in more detail, and promise to enhance our understanding of the EGG signal and vocal fold vibration.

ACKNOWLEDGMENTS

This research was supported by the Grant Agency of the Czech Republic, project GAČR 101/08/1155, and by the

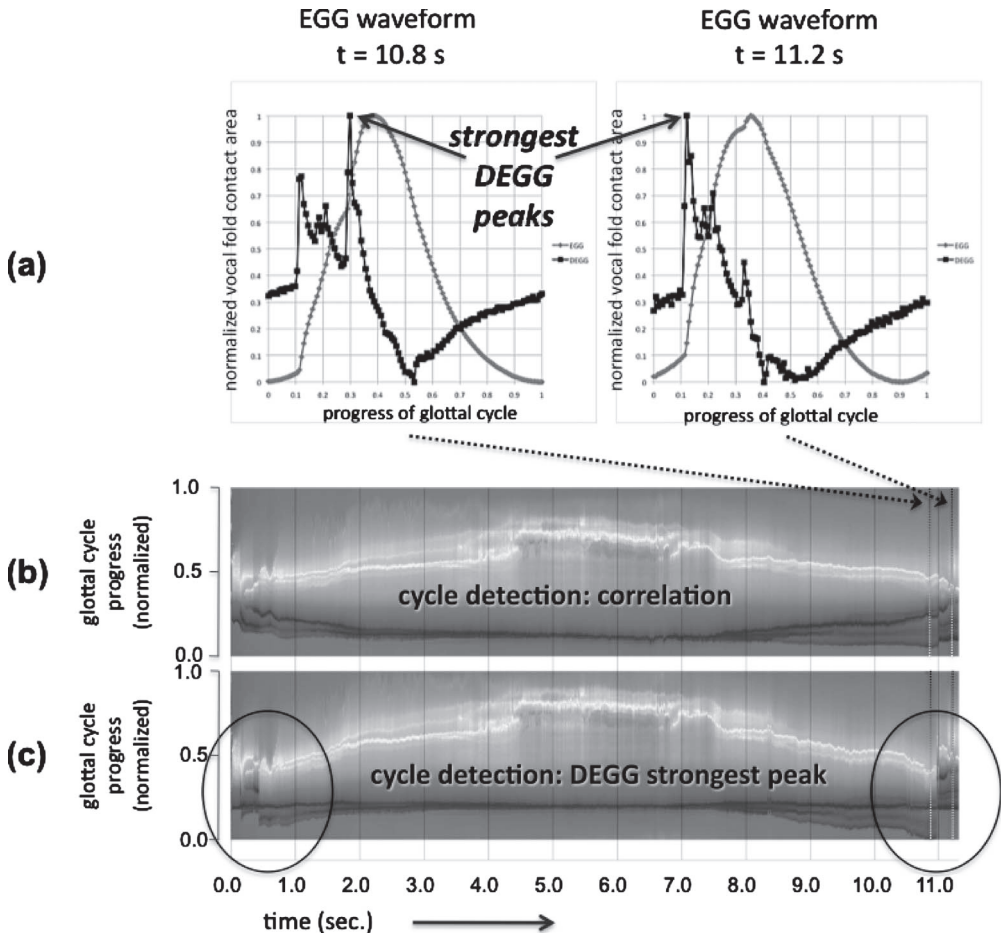


FIG. 10. Effects of different algorithms for glottal cycle detection when creating wavegrams. (a) EGG (black) and DEGG (gray) waveforms representing glottal cycles extracted at $t=10.8$ s, $t=11.2$ s; (b) DEGG wavegram, glottal cycle detection performed by correlating an ideal EGG waveform template with the analyzed EGG signal; (c) DEGG wavegram, glottal cycle detection performed by computing local positive maxima within the first derivative of the EGG signal (see Appendix for details).

ERC Advanced Grant "SOMACCA". We thank the anonymous reviewers for their insights and their helpful feedback.

APPENDIX

When creating a wavegram, the start and end of a glottal cycle can be chosen arbitrarily. In doing so, it is preferable that the entire contact phase lies within one cycle, i.e., the contact phase is not spread over two consecutive glottal cycles. To achieve this, two options have been implemented in the wavegram algorithm: (a) by correlating an ideal EGG waveform template with the analyzed EGG signal (as described above); or (b) by computing local positive maxima within the first derivative of the EGG signal, i.e., the DEGG signal. (The region of interest is one glottal period, the duration of which is known from F0 extraction.) The start of each glottal cycle is located at an arbitrary offset of the (normalized) glottal cycle duration, preceding each computed positive maximum in the DEGG signal. Offsets of 15% to 20% proved to give visually the best results, ensuring that even EGG waveforms with a very long contact phase (80% of the cycle duration) could be displayed in one entire glottal cycle.

The DEGG strongest peak alignment method provides the most intuitive information on the duration of the contact phase, since the contacting events within each glottal cycle (as indicated by the strongest positive maximum of the DEGG signal) are aligned to form a straight dark horizontal line in the wavegram (see Fig. 10(c)). This approach fails, however, if there are multiple positive peaks in the glottal contacting phase (i.e., the phase of a glottal cycle where the vocal fold area increases). This is particularly true if the amplitudes of the peaks vary (see for example the two EGG waveforms displayed in Fig. 8: at t=10.8 s the second DEGG peak is the strongest peak, and at t=11.2 s the first DEGG peak is stronger). In such a case, the alignment of the glottal cycles within the wavegram would abruptly change. Therefore, the "correlation method" has been chosen as the preferred default setting of the wavegram algorithm. In the current software, the "DEGG strongest peak method" has been implemented as a well, allowing us to explore both display options as needed.

Baer, T., Löfqvist, A., and McGarr, N. S. (**1983**). "Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques," J. Acoust. Soc. Am. **73**, 1304–1308.
Baken, R. J. (**1992**). "Electroglottography," J. Voice **6**, 98–110.
Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular, San Diego), pp. 413–427.
Berg, R., and Gall, V. (**1999a**). "Physiological and pathophysiological findings in glottal segmentation of voice-onset and voice-offset," in *Commu-*

nication and Its Disorders: A Science in Progress, Vol. 1, Proceedings of the 24th Congress International Association of Logopedics and Phoniatrics, Amsterdam, The Netherlands, edited by P. H. Dejonckere and H. F. M. Peters (International Association of Logopedics and Phoniatrics, Groningen, The Netherlands), pp. 9–12.

Berg, R., and Gall, V. (**1999b**). "Voice cascadation and electroglottography—Basics and practise," in *Communication and Its Disorders: A Science in Progress, Vol. 1*, Proceedings of the 24th Congress International Association of Logopedics and Phoniatrics, Amsterdam, The Netherlands, edited by P. H. Dejonckere and H. F. M. Peters (International Association of Logopedics and Phoniatrics, Groningen, The Netherlands), pp. 5–8.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in Proceedings of the Institute of Phonetic Sciences, Amsterdam, The Netherlands, pp. 97–110.

Childers, D., and Krishnamurthy, A. K. (**1985**). "A critical review of electroglottography," Crit. Rev. Biomed. Eng. **12**, 131–161.

Childers, D., Naik, J. M., Larar, J. N., Krishnamurthy, A. K., and Moore, G. P. (**1983**). "Electroglottography, speech, and ultra-high speed cinematography," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze (Denver Center for the Performing Arts, Denver, CO), pp. 203–220.

Childers, D. G., Hicks, D. M., Moore, G. P., and Alsaka, Y. A. (**1986**). "A model for vocal fold vibratory motion, contact area, and the electroglottogram," J. Acoust. Soc. Am. **80**, 1309–1320.

Colton, R. H., and Conture, E. G. (**1990**). "Problems and pitfalls of electroglottography," J. Voice **4**, 10–24.

Fabre, P. (**1957**). "Un procédé électrique percuntané d'inscription de l'accolement glottique au cours de la phonation: Glottographie de haute fréquence; premiers résultats (A non-invasive electric method for measuring glottal closure during phonation: High frequency glottography; first results)," Bull. Acad. Natl. Med. **141**, 66–69.

Fourcin, A., and Abberton, E. (**1971**). "First application of a new laryngograph," Med. Biol. Illus **21**, 172–182.

Fourcin, A., and Abberton, E. (**2008**). "Hearing and phonetic criteria in voice measurement: Clinical applications," Logoped. Phoniatr. Vocol. **33**, 35–48.

Fourcin, A., Abberton, E., Miller, D., and Howells, D. (**1995**). "Laryngograph: Speech pattern element tools for therapy, training and assessment," Eur. J. Disord. Commun. **30**, 101–115.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J. Acoust. Soc. Am. **115**, 1321–1332.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2005**). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency," J. Acoust. Soc. Am. **117**, 1417–1430.

Herbst, C., Howard, D. M., and Schlömicher-Thier, J. (**2010**). "Using electroglottographic real-time feedback to control posterior glottal adduction during phonation," J. Voice **24**, 72–85.

Herbst, C., and Ternström, S. (**2006**). "A comparison of different methods to measure the EGG contact quotient," Logoped. Phoniatr. Vocol. **31**, 126–138.

Hess, M. M., and Ludwigs, M. (**2000**). "Strobophotoglottographic transillumination as a method for the analysis of vocal fold vibration patterns," J. Voice **14**, 255–271.

Higgins, M. B., and Schulte, L. (**2002**). "Gender differences in vocal fold contact computed from electroglottographic signals: The influence of measurement criteria," J. Acoust. Soc. Am. **111**, 1865–1871.

Howard, D. M. (**1995**). "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," J. Voice **9**, 163–172.

Howard, D. M., Welch, G. F., Brereton, J., Himonides, E., DeCosta, M., Williams, J., and Howard, A. W. (**2004**). "WinSingad: A real-time display for the singing studio," Logoped. Phoniatr. Vocol. **29**, 135–144.

Kania, R. E., Hans, S., Hartl, D. M., Clement, P., Crevier-Buchman, L., and Brasnu, D. F. (**2004**). "Variability of electroglottographic glottal closed quotients: Necessity of standardization to obtain normative values," Arch. Otolaryngol. Head Neck Surg. **130**, 349–352.

Kitzing, P. (**1990**). "Clinical applications of electroglottography," J. Voice **4**, 238–249.

Koenig, W., Dunn, H. K., and Lacy, L. Y. (**1946**). "The sound spectrograph," J. Acoust. Soc. Am. **18**, 19–49.

Lohscheller, J., Eysholdt, U., Toy, H., and Döllinger, M. (**2008**). "Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics," IEEE Trans. Med. Imaging **27**, 300–309.

Miller, D. G., Švec, J. G., and Schutte, H. K. (**2002**). "Measurement of characteristic leap interval between chest and falsetto registers," J. Voice **16**, 8–19.

Miller, D. G., and Schutte, H. K. (**1999**). *The Use of the Electroglottograph in the Voice Studio*, in: Voice — Tradition and Technology. A state-of-the-art studio, edited by G. Nair (Singular, San Diego, CA), pp. 211–225.

Orlikoff, R. F. (**1991**). "Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects," J. Speech Hear. Res. **34**, 1066–1072.

Rossiter, D., Howard, D. M., and DeCosta, M. (**1996**). "Voice development under training with and without the influence of real-time visually presented biofeedback," J. Acoust. Soc. Am. **99**, 3253–3256.

Rothenberg, M. (**1979**). "Some relations between glottal air flow and vocal fold contact area," in Proceedings of the Conference on the Assessment of Vocal Pathology, ASHA Reports No. 11, edited by C. L. Ludlow and M. O. Hart (American Speech and Hearing Association, Rockville, MD), pp. 88–96.

Rothenberg, M., and Mahshie, J. J. (**1988**). "Monitoring vocal fold abduction through vocal fold contact area," J. Speech Hear. Res. **31**, 338–351.

Roubeau, B., Henrich, N., and Castellengo, M. (**2009**). "Laryngeal vibratory mechanisms: The notion of vocal register revisited," J. Voice **23**, 425–438.

Sapienza, C., Stathopoulos, E. T., and Dromey, C. (**1998**). "Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: Effects of measurement criteria and sound pressure level," J. Voice **12**, 31–43.

Scherer, R. C., Druker, D. G., and Titze, I. R. (**1988**). "Electroglottography and direct measurement of vocal fold contact area," in *Vocal Fold Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), Vol. **2**, pp. 279–290.

Schutte, H. K., and Miller, D. G. (**2001**). "Measurement of closed quotient in a female singing voice by electroglottography and videokymography," in Proceedings of the Fifth International Conference Advances in Quantitative Laryngology, Groningen, The Netherlands, edited by H. K. Schutte (Groningen Voice Research Lab, Groningen, The Netherlands).

Sedlacek, K., and Michek, F. (**1988**). "Klinische anwendung der dreidimensionalen periodizitätsanalyse (3D-PAN) [Clinical application of the three-dimensional period analysis (3D-PAN)]," in Proceedings of the 15th UEP Congress, Erlangen, Germany, pp. 11–13.

Smith, A., and Childers, D. (**1983**). "Laryngeal evaluation using features from speech and the electroglottograph," IEEE Trans. Biomed. Eng. **BME-30**, 755–759.

Švec, J. G., and Schutte, H. K. (**1996**). "Videokymography: High-speed line scanning of vocal fold vibration," J. Voice **10**, 201–205.

Švec, J. G., Sram, F., and Schutte, H. K. (**2007**). "Videokymography in voice disorders: What to look for?," Ann. Otol. Rhinol. Laryngol. **116**, 172–180.

Švec, J. G., Sundberg, J., and Hertegard, S. (**2008**). "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," J. Acoust. Soc. Am. **123**, 347–353.

Teaney, D., and Fourcin, A. (**1980**). "The electrolaryngograph as a clinical tool for the observation and analysis of vocal fold vibration," in Proceedings of the Ninth Symposium Care of the Professional Voice, New York, pp. 128–134.

Titze, I. R. (**1989**). "A four-parameter model of the glottis and vocal fold contact area," Speech Commun. **8**, 191–201.

Titze, I. R. (**1990**). "Interpretation of the electroglottographic signal," J. Voice **4**, 1–9.

Titze, I. R. (**1995**). "Workshop on acoustic voice analysis: Summary statement," National Center for Voice and Speech, Denver, CO, http://www.ncvs.org/museum-archive/sumstat.pdf (Last viewed 8/5/2010).

Vieira, M. N., McInnes, F., and Jack, M. A. (**1996**). "Analysis of the effects of electroglottographic baseline fluctuation on the F0 estimation in pathological voices," J. Acoust. Soc. Am. **99**, 3171–3178.

Vieira, M. N., McInnes, F., and Jack, M. A. (**2002**). "On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures," J. Acoust. Soc. Am. **111**, 1045–1055.

Supplement B: Manuscript II

Investigation of four distinct glottal configurations in classical singing - a pilot study

C. T. HERBST, S. TERNSTRÖM, J. G. ŠVEC

Reprinted with permission from
J Acoust Soc Am **125**, EL104-EL109 (**2009**)
 [DOI: 10.1121/1.3057860]

Investigation of four distinct glottal configurations in classical singing—A pilot study

Christian T. Herbst
Laboratory of Biophysics, Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, tř. Svobody 26, 77146 Olomouc, Czech Republic, and Tölzer Knabenchor, Herbert-von-Karajan-Platz 2, 5020 Salzburg, Austria
herbst@ccrma.stanford.edu

Sten Ternström
Department of Speech, Music and Hearing, School of Computer Science and Communication, Kungliga Tekniska Högskolan, Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden
stern@kth.se

Jan G. Švec
Laboratory of Biophysics, Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, tř. Svobody 26, 77146 Olomouc, Czech Republic
svecjan@vol.cz

Abstract: This study investigates four qualities of singing voice in a classically trained baritone: "naïve falsetto," "countertenor falsetto," "lyrical chest" and "full chest." Laryngeal configuration and vocal fold behavior in these qualities were studied using laryngeal videostroboscopy, videokymography, electroglottography, and sound spectrography. The data suggest that the four voice qualities were produced by independently manipulating mainly two laryngeal parameters: (1) the adduction of the arytenoid cartilages and (2) the thickening of the vocal folds. An independent control of the posterior adductory muscles versus the vocalis muscle is considered to be the physiological basis for achieving these singing voice qualities.

PACS numbers: 43.75.Rs, 43.70.Gr [TR]
Date Received: June 26, 2008 **Date Accepted:** November 25, 2008

1. Introduction

The ability to control voice quality in singing is crucial for a singer, yet the information on the specific mechanisms used in singing has been unsatisfactory. Probably the most controversial topic of singing voice quality from the historical perspective has been that of voice registers (Titze, 2000). The two main registers of the singing voice, particularly in males, have been recognized to be the chest and falsetto register. Most singers, however, are capable of blending the registers and produce more than just two distinct voice qualities. Information on how these qualities can be achieved has been largely insufficient.

This study investigates a classically trained baritone who demonstrated the ability to produce four distinct voice qualities, which he considered to be the "building blocks" of his singing technique. He called those four qualities as Type A "*Naïve singer's falsetto*," Type B "*Countertenor falsetto*," Type C "*Lyrical style*," and Type D "*Full chest*." Acoustic recordings of these four phonation types are demonstrated in the audio file Mm. 1. Based on his self-perception, the subject claimed that these four voice qualities are created with four different laryngeal configurations. The specific goal of this study was to investigate these phonations and the laryngeal adjustments in these four phonation types. The more general goal was to establish better understanding of the laryngeal adjustment strategies that are used to control the voice quality in singing.

[Mm. 1. Acoustic recordings of the four phonation types A–D (naïve falsetto, counter tenor falsetto, lyrical chest, full chest). This is a file of type .wav (1.1 MB).]

2. Material and methods

The investigated subject was a baritone with a university degree in voice pedagogy and 15 years of experience in classical singing (the author C.H.). He produced eight to ten sustained phonations in each of the four phonation qualities at a fundamental frequency of 294 Hz (tone D4). This frequency was at the second passaggio of the baritone, where both chest and falsetto phonations were possible. Vowel /i/, which allows examination through rigid laryngoscopy, was chosen for all phonatory tasks.

The adjustment and vibration of the vocal folds was observed with a rigid endoscope using two alternative techniques—laryngeal videostroboscopy (Bless *et al.*, 1987; Baken and Orlikoff, 2000) and videokymography (Švec and Schutte, 1996). The specific audiovisual equipment used was identical to the one used in the study of Švec *et al.* (2008). The microphone and electroglottographic signals were recorded simultaneously with the video signals and stored in the two audio channels of the final digital video file. The video signals were digitized using a Pinnacle DV 500 System set to full PAL mode (25 images/s, 720 × 525 pixels resolution) and stored as AVI files using the Intel Indeo 5.0 codec.

The subject repetitively produced each phonation type while the position and focus of the endoscope was adjusted to obtain videokymographic images of the best quality at the place of maximum vibration amplitude of the vocal folds. After satisfactory videokymographic images were obtained from each phonation types, the videokymographic camera was replaced by the standard color camera and the continuous light was replaced by stroboscopic light. Then the subject repeated the phonations for all the types and strobovideolaryngoscopic recordings were done.

From laryngeal videostroboscopy two images were extracted, both at maximum glottal closure and maximum glottal opening to document the laryngeal configuration as well as the vibration mode of the vocal folds. From videokymographic recordings, representative phonations were extracted from the stable portion in the middle of each video sample; no voicing onsets and offsets were included in the analyzed data. For the sustained notes sung in the four distinct phonation types, the following criteria had to apply for data to be selected for further evaluation: (1) Phonation was stable. (2) The minimum duration exceeded one second. (3) The subject confirmed that in his opinion the phonation did represent the attempted phonation type. (4) Videokymographic images showed sharp and clear glottal contours and the scan line was located approximately at the place of the maximum vibration amplitude of the vocal folds. In this way three to five representative video samples per phonation type were selected out of a total of 115 videokymographically recorded phonations, resulting in a total of 15 samples.

The videokymographic closed quotient (CQ_{VKG}) was determined using the formula $CQ_{VKG}=t_c/T_0$ where t_c is the duration of the closed phase and T_0 is the duration of the vibratory cycle. The durations were measured by manually counting the number of pixels for open and closed phase. The pixel spacing corresponded to a time interval of 128 μs, or about 26.5 pixels per period of the fundamental frequency. For each of the 15 samples of stable phonation, the CQ_{VKG} was calculated from four consecutive glottal cycles at two or three (in short or long recordings, respectively) different instances equally distributed over the duration of the phonatory sample. The data points were separated by at least eight video frames (320 ms). CQ_{VKG} readings were then averaged per phonation type.

The microphone signal was low-pass filtered with a cutoff-frequency of 8 kHz (to remove an artifact noise around 11 kHz produced by the light source during videokymographic recordings) and the sampling rate was reduced to 16 kHz. The alpha ratio was calculated as the ratio of the high-band energy (1000–5000 Hz) and the low-band energy (up to 1000 Hz) of the acoustic signal, expressed in dB (Löfqvist and Mandersson, 1987). Typical alpha ratio values are negative numbers, which increase (i.e., become less negative) as the spectrum of the signal becomes flatter (i.e., the signal has stronger high-frequency components).

The electroglottographic signal was used to obtain the electroglottographic contact quotient (CQ_{EGG}), which was calculated by a criterion level method with a threshold level of 25% as has been used, e.g., by Orlikoff (1991) or Herbst and Ternström (2006). Since the elec-

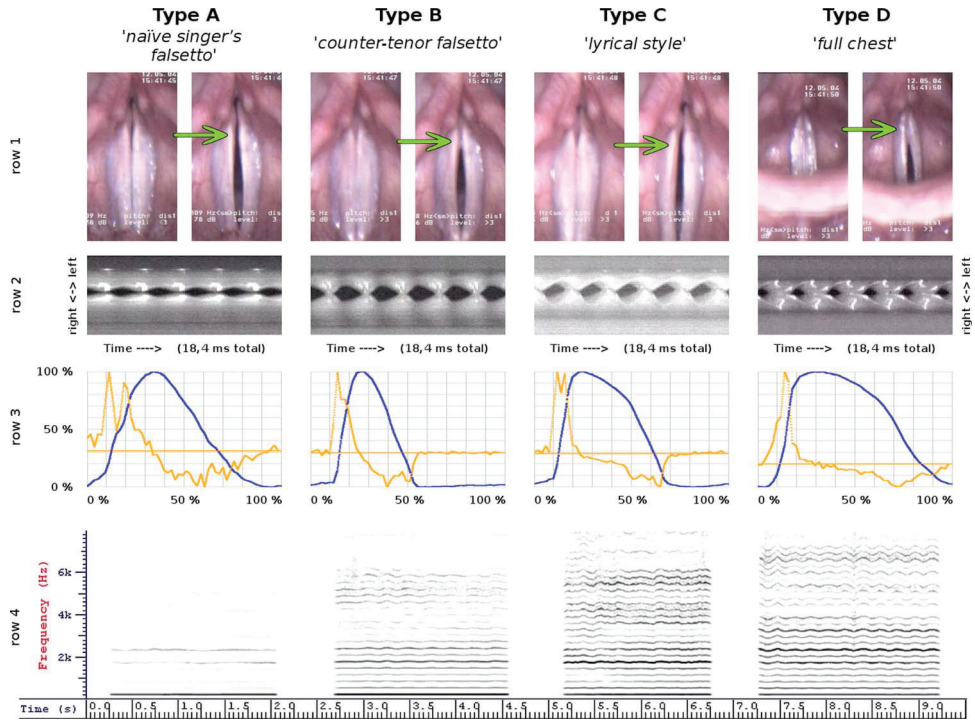


Fig. 1. (Color online) Images and signals documenting the vocal fold vibration in the four phonation types A–D. From top to bottom: Row 1: Pairs of videostroboscopic images at the phases of maximum vocal fold contact and maximum glottal opening. The arrows point at the position of the vocal processes of the arytenoid cartilages which are apart in types A and C and closed in types B and D. Row 2: Typical videokymographic images at the place of maximal vibration amplitude of the vocal folds. Row 3: Typical EGG signals (dark/blue) and their first derivative (light/orange online) for phonation types A–D, normalized both in amplitude and time. The *x*-axis represents normalized glottal cycle duration, and the *y*-axis shows EGG and DEGG signal amplitude, locally normalized per glottal cycle. Row 4: Spectrogram of sustained notes (taken from the audio channel of the videokymographic recordings) produced with phonation types A, B, C, and D consecutively. The videostroboscopic recordings of the four phonation types are documented in the media file below.

troglottograph used had no automatic gain controller, it was also possible to evaluate the strength of the EGG signal, which was quantified by computing the time-varying rms value of the signal, using a window duration of 125 ms. The values were expressed on a dB scale with an arbitrary reference value.

3. Results

Examples of the videostroboscopic recordings of the four phonation types are given in the video file Mm. 2.

[Mm. 2. Videostroboscopic recordings of the four phonation types A–D (naïve falsetto, counter tenor falsetto, lyrical chest, full chest). This is a file of type .avi (7.4 MB).]

The laryngeal adjustments in the four types of phonations are shown in the top section of Fig. 1. The images reveal distinct adjustments of the laryngeal structures and vibratory features of the vocal folds for each phonation type. The most distinct differences were seen (1) at the posterior, cartilaginous part of glottis, which was varying between slightly open (types A and C) and closed (types B and D); (2) at the vocal processes of the arytenoid cartilages (marked by arrows in the stroboscopic images of Fig. 1, row 1), which were in some cases vibrating with the vocal folds (types A and C) and in other cases pressed together and not vibrating (type D; in phonation type B, the vocal processes were intermittently pressed together

Table 1. Glottal configurations and vibratory features, as seen in laryngeal videostroboscopy for phonation types A–D.

	Type A ("naïve singer's falsetto")	Type B ("countertenor falsetto")	Type C ("lyrical style")	Type D ("full chest")
Posterior glottis	slightly abducted	closed	barely closed	closed
Vocal processes vibration	yes[a]	intermittent[b]	yes[a]	no[c]
Mucosal waves extent	very short[d]	very short[d]	medium	long

[a]Vibrating together with the focal folds.
[b]Vocal processes alternated between slightly vibrating and pressed together.
[c]Vocal processes pressed together.
[d]Seen only in posterior part of vocal folds.

or vibrating with the vocal folds); and (3) in the mucosal waves on the vocal folds, the extent of which varied from long (type D) to very short (types A and B). The specific findings for each of the four phonation types are summarized in Table 1.

The videokymographic images revealed distinct differences in the vibratory pattern of the vocal folds among the four phonation types (Fig. 1, row 2). Considering the categorization of vibration characteristics as specified by Švec *et al.* (2007), the most prominent differences were recognized in (1) duration of closed phase, which increased when going from type A to type D; (2) roundedness of the lateral peaks, which successively sharpened when going from type A to type D; and (3) the extent of laterally travelling mucosal waves (only short mucosal waves for types A and B in contrast to medium/long mucosal waves for types C and D, respectively). The specific findings for each of the four phonation types are summarized in Table 2.

For acoustic analysis, the audio signal was extracted from the same 15 video samples that were considered for CQ_{VKG} analysis. Spectral analysis showed an increasing energy content in high frequency partials from type A to D. For phonation types A, B, and C this corresponded to a steady increase of alpha ratio values (Table 3). The types C and D showed similar alpha ratios but different distribution of spectral maxima: type C showed more energy in the region of 3500–6000 Hz, whereas the D type had more energy in the region of 2300 to 3500 Hz (see Fig. 1, row 4).

The EGG signals showed an increase of vocal fold contact duration when going from type B through C to D. This was reflected in the increase of the calculated average CQ_{EGG} (Table 3). The EGG waveform for phonation type A had a much smaller amplitude than the other types

Table 2. Videokymographic findings for phonation types A–D. For closed quotient mean values and standard deviations are listed; the numbers in the parentheses give the ranges of the values observed.

Phonation type	No. of phonations/ measurements	Closed quotient (CQ_{VKG})	Lateral peaks shape	Mucosal waves extent
A	3/32	**0**	rounded	none/barely visible
B	4/40	**0.28** ± 0.05 *[0.2, 0.37]*	rather rounded	none/barely visible
C	3/36	**0.47** ± 0.05 *[0.37, 0.57]*	rather sharp	medium
D	5/60	**0.69** ± 0.04 *[0.63, 0.78]*	sharp	long

Table 3. Acoustic and electroglottographic signal characteristics for phonation types A–D: alpha ratio, SPL values, CQ_{EGG}, and averaged EGG signal strength (expressed in dB with arbitrary reference). All computed data are indicated as mean values with standard deviation and the extreme values in the parentheses.

		Acoustic signal		Electroglottography	
Phonation type	No. of phonations	Alpha ratio (dB)	SPL@30 cm (dB *re* 20 μPa)	EGG signal strength (dB)	CQ_{EGG}
A	3	**−16.3** ± 0.8 *[−16.78, −15.46]*	**80** ± 1.8 *[77, 81]*	**−26** ± 1.3 *[−27, −24]*	**N/A**[a]
B	4	**−9.2** ± 1.6 *[−10.4, −6.3]*	**88** ± 0.9 *[87, 89]*	**−18** ± 0.3 *[−18, −18]*	**0.31** ± 0.01 *[0.30, 0.33]*
C	3	**0.3** ± 1.6 *[−1.4, 1.7]*	**84** ± 2.2 *[81, 86]*	**−16** ± 0.6 *[−17, −15]*	**0.48** ± 0.07 *[0.40, 0.59]*
D	5	**−1.1** ± 1.7 *[−2.5, 1.9]*	**82** ± 2.1 *[79, 84]*	**−17** ± 1.7 *[−20, −15]*	**0.65** ± 0.01 *[0.64, 0.66]*

[a]See text.

and a quasi-sinusoidal shape, which suggested that there was no full contact of the vocal folds. Thus, we concluded, in accordance with Herbst and Ternström (2006), that it is not appropriate to calculate a CQ_{EGG} based on that kind of signal.

4. Discussion

The data supported the original hypothesis that the four singing types are produced with four different laryngeal adjustments. The differences were directly observed with VKG and laryngeal videostroboscopy and were also reflected in the electroglottographic and microphone signals. The most remarkable factor for distinguishing the types A from B and C from D was found to be the configuration of the posterior glottis and the position of the vocal processes. In types B and D the posterior glottis was fully adducted and the vocal processes were mostly pressed together, thus actively shortening the vibrating part of the vocal folds to only the membranous part. On the other hand, in the A and C types the posterior glottis was slightly abducted and the vocal processes were participating in the vibration of the vocal folds (Table 1).

The singer indicated that the phonation types A and B were produced in falsetto voice whereas types C and D were of chest register quality. It has been known that chest versus falsetto control is physiologically achieved mostly by the vocalis muscle (Hirano, 1974; Titze, 2000). Activity of the vocalis muscle thickens the body of the vocal fold and slackens the vocal fold cover, resulting in vibration with larger vertical phase differences and more pronounced mucosal waves (Hirano, 1974; Titze, 2000). Generally, these characteristics are expected to be reflected in the increased sharpness of the lateral peaks in the videokymographic images of the vocal folds, as well as in the increased extent of the mucosal waves on the upper vocal fold surface (Švec *et al.*, 2008). Indeed, the videokymographic images of type A and B phonations show more rounded lateral peaks and shorter laterally traveling mucosal waves than the types C and D (Table 2) thus objectively supporting the singer's assumption that types A and B belong rather to falsetto register, whereas types C and D rather belong to chest register.

Based on these findings, it may be appropriate to call the phonation types A and B "abducted falsetto" and "adducted falsetto" [referred to as "open-chink falsetto" and "closed-chink falsetto" by Rubin and Hirt (1960)], whereas the types C and D can be seen as "abducted chest" and "adducted chest" registers, respectively. The differences in the closed phase seen in Table 2 can be explained as a consequence of these adjustments.

While register control is physiologically achieved mostly by the vocalis muscle (Hirano, 1974; Titze, 2000), posterior glottal adduction is known to be regulated by laryngeal adductors, such as posterior cricoarytenoid (PCA) and interarytenoid (IA) muscles (Titze, 2000). The present results suggest that in singing, the adduction of vocal processes is actively

used as a parameter in voice quality control, which is independent and separate from that of the (chest-falsetto) register control by the vocalis muscle. While the posterior glottal adduction has been recognized as an independent parameter in voice production (e.g., Titze, 2000), the use of this parameter separately from the chest-falsetto control has not, to our knowledge, been documented in singing before. An independent control and flexibility of the adductory muscles appears to be an important factor in allowing the production of different singing qualities.

This pilot study is limited in that it documents the laryngeal behavior laryngoscopically in only a single subject. However, it is not too far-fetched to conjecture that the independent control of the vocal fold body tension (through vocalis muscle) and posterior glottal adduction (through PCA and IA muscles) plays an important role in controlling singing voice quality in general. In classical music different voice qualities are expected for modulating the timbre within a musical piece, in order to enhance the expression of the artistic performance. It is conceivable that the lack of independent control of these two parameters may lead to problems in singing students who fail in singing specific styles. In future studies we plan to address this hypothesis further using a larger number of singer subjects.

5. Conclusions

The results of this study showed that the four phonation types were acoustically distinct and that they were indeed produced with different laryngeal settings. These settings could be explained by the independent manipulation of mainly two laryngeal parameters: (1) *the thickening of the vocal folds* and (2) *the adduction of the posterior glottis*. These two physiologic parameters represent two physiologically distinct types of glottal adduction: membranous adduction (adjustable by thyroarytenoid-vocalis muscles) and cartilaginous adduction (adjustable by cricoarytenoid and interarytenoid muscles). The two types of glottal adduction should be separated from each other when studying different voice qualities in singing.

Acknowledgments

The data for this study were obtained in 2004 during the authors' stays at the Department of Speech, Music and Hearing, KTH, Stockholm. C. Herbst's stay was supported by the Erasmus Student Exchange Programme of the European Commission, and J. Švec's stay was supported by an individual grant from the Wenner-Gren Foundation. In 2008, the study was supported by the Grant Agency of the Czech Republic, Project GAČR 101/08/1155. The authors thank Hans Larsson at the Department of Logopedics and Phoniatrics, Karolinska University Hospital at Huddinge, Stockholm, for his help in acquiring the laryngoscopic recordings.

References and links

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Publishing Group, San Diego, CA).
Bless, D. M., Hirano, M., and Feder, R. J. (**1987**). "Videostroboscopic Evaluation of the Larynx," Ear Nose Throat J. **66**(7), 289–296.
Herbst, C., and Ternström, S. (**2006**). "A Comparison of Different Methods to Measure the EGG Contact Quotient," Logoped. Phoniatr. Vocol. **31**(3), 126–138.
Hirano, M. (**1974**). "Morphological Structure of the Vocal Cord as a Vibrator and its Variations," Folia Phoniatr. **26**(2), 89–94.
Löfqvist, A., and Mandersson, B. (**1987**). "Long-time average spectrum of speech and voice analysis," Folia Phoniatr. **39**(5), 221–229.
Orlikoff, R. F. (**1991**). "Assessment of the Dynamics of Vocal Fold Contact From the Electroglottogram: Data From Normal Male Subjects," J. Voice **34**(5), 1066–1072.
Rubin, H. J., and Hirt, C. C. (**1960**). "The falsetto: A high speed cinematographic study," Laryngoscope **70**, 1305–1324.
Švec, J., and Schutte, H. K. (**1996**). "Videokymography: High-speed line scanning of vocal fold vibration," J. Voice **10**(2), 201–205.
Švec, J., Šram, F., and Schutte, H. K. (**2007**). "Videokymography in Voice Disorders: What to Look For?," Ann. Otol. Rhinol. Laryngol. **116**(3), 172–180.
Švec, J., Sundberg, J., and Hertegård, S. (**2008**). "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," J. Acoust. Soc. Am. **123**(1), 347–353.
Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Iowa City, IA).

Supplement C: Manuscript III

Membranous and cartilaginous vocal fold adduction in singing

C. T. HERBST, Q. QIU, H. K. SCHUTTE, J. G. ŠVEC

Reprinted with permission from
J Acoust Soc Am **129**, 2253-2262 (**2011**)
 [DOI: 10.1121/1.3552874]

Membranous and cartilaginous vocal fold adduction in singing[a)]

Christian T. Herbst[b)]
Department of Biophysics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

Qingjun Qiu and Harm K. Schutte
Cymo B.V., Stavangerweg 21-2, 9723 JC, Groningen, The Netherlands

Jan G. Švec
Department of Biophysics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

(Received 14 January 2010; revised 5 January 2011; accepted 5 January 2011)

While vocal fold adduction is an important parameter in speech, relatively little has been known on the adjustment of the vocal fold adduction in singing. This study investigates the possibility of separate adjustments of cartilaginous and membranous vocal fold adduction in singing. Six female and seven male subjects, singers and non-singers, were asked to imitate an instructor in producing four phonation types: "aBducted falsetto" (FaB), "aDducted falsetto" (FaD), "aBducted Chest" (CaB), and "aDducted Chest" (CaD). The phonations were evaluated using videostroboscopy, videokymography (VKG), electroglottography (EGG), and audio recordings. All the subjects showed less posterior (cartilaginous) vocal fold adduction in phonation types FaB and CaB than in FaD and CaD, and less membranous vocal fold adduction (smaller closed quotient) in FaB and FaD than in CaB and CaD. The findings indicate that the exercises enabled the singers to separately manipulate (a) *cartilaginous adduction* and (b) *membranous medialization* of the glottis though vocal fold bulging. Membranous adduction (monitored via videokymographic closed quotient) was influenced by both membranous medialization and cartilaginous adduction. Individual control over these types of vocal fold adjustments allows singers to create different vocal timbres.
 [DOI: 10.1121/1.3552874]

PACS number(s): 43.75.Rs, 43.70.Gr [DAB] Pages: 2253–2262

I. INTRODUCTION

In both classical and commercial contemporary music, different voice qualities are expected within a musical piece, in order to enhance the expression of the artistic performance. Apart from changing the shape of the vocal tract (Echternach *et al.*, 2008, 2010; Gullaer *et al.*, 2006; Henrich *et al.*, 2007; Joliveau *et al.*, 2004; Schutte *et al.*, 2005; Story, 2004; Sundberg, 1972, 1974; Sundberg and Skoog, 1997; Wendler, 2008), the voice quality can be considerably influenced by changing the vibration pattern of the vocal folds (Henrich *et al.*, 2005; Salomão and Sundberg, 2009; Sundberg and Högset, 2001; Švec *et al.*, 2008). Here, one of the most critical factors is the vocal fold aDduction; i.e. bringing the vocal folds together (as an opposite to aBduction, i.e., taking the vocal folds apart).

Clinically, it has been known that inadequate vocal fold adduction is a frequent cause of voice problems. In singing, vocal fold aDduction/aBduction has been recognized to play an important role for the production of voice registers. For instance, Titze (2000) defined an "abduction quotient" as a ratio between the pre-phonatory glottal half width and the amplitude of vocal fold vibration, and related it to the production of vocal registers. Most frequently, the vocal fold adduction has been quantified in singing indirectly via the so-called "open quotient" (OQ), quantifying the relative portion of time the vocal folds are apart within their vibration period; or its reciprocal "closed quotient" (CQ), quantifying the relative portion of time the vocal folds are closed within their vibration period (Baken and Orlikoff, 2000; Barlow and Howard, 2005; Henrich *et al.*, 2004; Hirano, 1981; Moore and von Leden, 1958; Sundberg *et al.*, 2005; Timcke *et al.*, 1958). A sudden decrease of the CQ has been reported, e.g., during the transition from chest/modal to falsetto register (Henrich *et al.*, 2005; Miller *et al.*, 2002; Švec *et al.*, 2008).

Along their length, the vocal folds can be divided into the *membranous portion*, i.e. the part between the anterior commissure and the vocal processes of the arytenoid cartilages; and the *cartilaginous portion*, i.e. the portion between the vocal processes and the posterior commissure of the vocal folds (e.g., Titze, 1989). The cartilaginous portion is adducted mostly via two main adductory muscles: lateral cricoarytenoid (LCA) bringing the vocal processes (i.e., anterior prominences of the arytenoid cartilages) together; and interarytenoid (IA) muscles bringing the posterior parts of the arytenoid cartilages together (Baken and Isshiki, 1977; Berg van den *et al.*, 1960; Broad, 1968; Fried *et al.*, 2009; Hunter *et al.*, 2004; Letson and Tatchell, 1997; Sataloff,

[a)]This study was presented at the 8th Pan-European Voice Conference 2009 in Dresden, Germany.
[b)]Author to whom correspondence should be addressed. Electronic mail: herbst@ccrma.stanford.edu

1997; Zemlin, 1997). An antagonistic aBductory function is provided mainly by the posterior cricoarytenoid muscles (PCA) which move the vocal processes apart, and therefore play a crucial role in enabling breathing. A supplementary role in arytenoids adduction/abduction is played also by the thyroarytenoid (TA) and cricothyroid (CT) muscles, for details see Zemlin (1997). An incomplete adduction of the cartilaginous portion of the vocal folds can be recognized laryngoscopically as a "posterior glottal chink" (PGC) (e.g., Södersten *et al*., 1995). Variation of the cartilaginous adduction has been recognized to play an important role in speech—in the production of breathy, normal, and pressed or creaky voice (Ladefoged, 1975; Zemlin, 1997). Whereas in breathy voice the arytenoid cartilages are set apart, in pressed voice they are usually squeezed together.

The cartilaginous adduction, specifically the adduction of the vocal processes of the arytenoid cartilages, directly influences the position of the membranous portion of the vocal fold. However, the membranous vocal fold can additionally be adducted also via the activity of the TA muscle by bulging the body of the vocal fold. This process can be labelled as "*membranous medialization*,"since it shifts the membranous portion of the vocal folds in medial direction. The activity of the TA muscle and the related changes in vocal fold geometry were recognized to play an important role for voice registers in singing (Choi *et al*., 1993; Hirano, 1974; Titze, 2000).

In a previous case study, we observed that the cartilaginous and membranous adduction of the vocal folds may be adjusted separately for the purpose of creating different sound qualities in singing. Particularly, the membranous adduction was found to play a role in distinguishing between the chest and falsetto registers, whereas the cartilaginous adduction played a role in distinguishing between "lyrical chest" and "heavy chest" register as well as in distinguishing between the "naïve" and "counter-tenor falsetto" produced by a single trained baritone singer (Herbst *et al*., 2009). The goal of this study was to explore the possibilities of a separate adjustment of the cartilaginous and membranous vocal fold adduction in a number of singers and non-singers.

II. METHODS

A. Design of phonatory exercises

At first, singing exercises were designed, which targeted the separate manipulation of cartilaginous and membranous vocal fold adduction. The four singing voice qualities used previously by Herbst *et al*. (2009) were taken as a basis for the investigations here. The chest vs falsetto registers were used as the putative means to manipulate the membranous vocal fold adduction. The degree of breathiness, on the other hand, was used as the putative means to manipulate the posterior vocal fold adduction. The resulting four phonation qualities are labeled here as "aBducted falsetto" (FaB), "aDducted falsetto" (FaD), "aBducted chest" (CaB), and "aDducted chest" (CaD).

B. Subjects and tasks

Six female and seven male singers and non-singers, according to the categorization of Bunch and Chapman (2000), participated in the experiment (see Table I). All subjects participated in this study voluntarily. They received a 30 min training session, in which they were presented with the four phonation types (i.e., FaB, FaD, CaB, and CaD) as demonstrated by an instructor (author CH, who was the subject in the previous pilot study). Acoustic and laryngoscopic samples of those four phonation types are freely available as an attachment to the previous publication online (Herbst *et al*., 2009).

The singers were asked to vocally imitate the instructor until a consensus was reached between the instructor and individual subjects that they did achieve the targeted phonation type. In case the singers (the less experienced ones) were not succeeding in imitating the phonations, they were given the additional instruction to sing either more "breathy"

TABLE I. Demographic data of all subjects, singer categorization (Bunch and Chapman, 2000) and assigned target notes for both falsetto and chest phonations. The singer categories used are: 3.9: National/Big City–Cabaret and Club; 4.1a: Regional/Touring–Opera; 4.5: Regional/Touring–Concert/Oratorio/Recital; 5.4: Local/Community–Concert/Oratorio/Recital; 5.5: Local/Community–Church Soloist; 8: Amateur (sings for pleasure).

Initials	Age	Sex	Voice class	Years of training	Years of experience	Category	Additional instruction ("breathy," "pressed")	Target note falsetto	Target note chest
RW	33	f	Soprano	15	13	4.5	No	Eb4	Eb4
HM	35	f	Mezzo-Soprano	0	0	8	Yes	F4	D4
ME	58	f	Mezzo-Soprano	8	0	8	Yes	D4	D4
WB	53	f	Mezzo-Soprano	8	12	5.4	No	D4	D4
MM	56	f	Alto	24	20	3.9	No	D4	D4
KP	26	f	N/A[a]	0	0	N/A[a]	Yes	E4	Eb4
HS	63	m	Tenor	1	0	8	Yes	D4	D4
CH	35	m	Baritone	10	3	5.5	N/A (instructor)	D4	D4
JS	39	m	Baritone	0	0	8	No	E4	E4
MD	25	m	Baritone	7	2	8	No	D4	D4
MK	25	m	Baritone	0	0	8	yes	E4	E4
DM	73	m	Bass	45	45	4.1a	No	C#4	C#4
QQ	32	m	N/A[a]	0	0	N/A[a]	Yes	Eb4	C4

[a]Subjects KP and QQ considered themselves non-singers (i.e., not even amateur singers) and are therefore not covered by the Bunch and Chapman (2000) taxonomy.

("aBducted falsetto" and "aBducted chest") or more "pressed" ("aDducted falsetto" and "aDducted chest"). An overview as to which subject did receive this additional instruction is given in Table I.

During the training session, each subject's transition region (zona di passaggio) was established as the range of pitches at which the target notes of all the four phonation types could be reached. When possible, identical pitches were chosen for target notes of phonations in both falsetto and chest registers. In order to make the desired registration (chest or falsetto) easier, the target notes were reached by singing a descending (for falsetto) or ascending (for chest) scale of five notes. The subjects were asked not to "blend or mix the registers." The vowel /i/, which allows examination through rigid laryngoscopy, was used for all phonatory tasks.

After completion of the training session, the subjects were asked to produce the four targeted phonation types during simultaneous capture of acoustic and electroglottographic data and laryngeal imaging. All the phonatory tasks were repeated four times and were recorded with (a) an audio recording equipment, without laryngeal imaging; (b) videokymography; (c) videolaryngoscopy; and (d) videostrobolaryngoscopy. For videostrobolaryngoscopic recordings, stable phonation on each pitch had to be sustained for about 2 s, in order to allow for at least four complete vibratory cycles of the vocal folds to be captured.

C. Equipment setup

All recordings took place in a laboratory room at the Department of Biomedical Engineering of the University Groningen, the Netherlands. Acoustic data were captured with an omnidirectional head-mounted microphone (type AKG C417 PP, AKG Acoustics GmbH, Vienna, Austria). The microphone was mounted at a spectacle frame (without glasses) worn by the subject. The microphone was attached at a distance of c. 7 cm and at an angle of c. 45° horizontally to the subject's mouth. The microphone signal was amplified with a Behringer Tube Ultragain MIC100 pre-amplifier (BEHRINGER International GmbH, Willich, Germany). SPL calibration was performed at a distance of 30 cm with a Brüel and Kjaer measuring amplifier type 2609 (Brüel Kjaer Sound & Vibration Measurement A/S, Naerum, Denmark) and a Brüel and Kjaer 4132 (Brüel Kjaer Sound & Vibration Measurement A/S, Naerum, Denmark) condenser microphone. Due to a system's malfunction, no microphone data could be recorded for subjects DM and WB.

All laryngeal imaging was performed with a Wolf 4450.7 90° rigid endoscope (Richard Wolf GmbH, Knittlingen, Germany). Videolaryngoscopic and videostroboscopic data were recorded with an Alphatron Stroboview 2000 electronic videostroboscopic system (Alphatron Medical Systems BV, Rotterdam, The Netherlands). Videokymographic (VKG) images were captured with a VKG 2 cymo prototype camera (Cymo B.V., Groningen, The Netherlands) (Qiu and Schutte, 2006), the R.Wolf 5261.27 c-mount lens adapter (Richard Wolf GmbH, Knittlingen, Germany), and the KAY 300 W xenon light source model 7150 (KayPENTAX, Lincoln Park, NJ).

The audio and video data were recorded using a Panasonic NV-GX7 mini DV camera (Panasonic Corporation of North America, Secaucus, NJ) using an analogue audio/video input. Afterwards, the data were imported to a PC via a firewire connection and saved as individual files in AVI format. The electroglottographic signal was captured with a Glottal Enterprises EGG2-PC electroglottograph (Glottal Enterprises, Syracuse, New York) and monitored with a Tektronix TDS 210 oscilloscope (Tektronix Inc., Beaverton, OR).

D. Data analysis

For the tones sung in the four distinct phonation types, the following criteria were applied to the data selected for further evaluation: (1) the minimum tone duration exceeded more than 1 s. (2) VKG images showed sharp and clear glottal contours and the scan line was located approximately at the place of the maximum vibration amplitude of the vocal folds, perpendicularly to the glottal axis. (3) Analysis of spectrographic data and inspection of electroglottographic waveforms did not show abrupt (involuntary) changes when singing the ascending (for chest register) or descending (for falsetto register) scale before arriving at the target note. In this way, one representative sample for each subject and each phonation type was selected from the total of 776 samples. In the case of the FaD phonations of subjects ME and KP, the qualities of the criteria were not met. Therefore, the phonations of the two semi-tones above the target note were chosen in these specific cases for further analysis.

The cartilaginous adduction was evaluated by measuring the posterior glottal chink (PGC). For each phonation type of each subject, representative images taken at the moment of maximum glottal closure were extracted from the videostroboscopic data. Each extracted image was converted to 16-bit greyscale color, and was enlarged by a factor of 8 with the image processing software FIJI/ImageJ. In order to better detect the contours of the glottis, the image contrast was maximized by running the FIJI command "enhance contrast" with the options "normalize" and "equalize histogram" enabled. The glottal chink was defined as the (laryngoscopically visible) space between the arytenoid cartilages, delimited anteriorly by the tip of the vocal process. This region, expressed in pixels, was manually measured with the FIJI/ImageJ image processing software. The glottal chink areas were averaged per subject over the four phonation types, and each individual glottal chink area was divided by this average as a means of intra-subject data normalization.

The membranous glottal adduction was estimated indirectly by measuring the videokymographic closed quotient CQ_{VKG} at the middle of the membranous portion of the glottis (Herbst and Ternström, 2006; Qiu *et al.*, 2003). The CQ_{VKG} was determined for the representative samples using the formula $CQ_{VKG} = t_c/T_0$ where t_c is the duration of the closed phase and T_0 is the duration of the vibratory cycle. The durations were measured manually by counting the number of pixels for open and closed phases. The pixel spacing corresponded to a time interval of 128 μs, providing about 22,3–28,2 pixels (depending on target note) per period of the

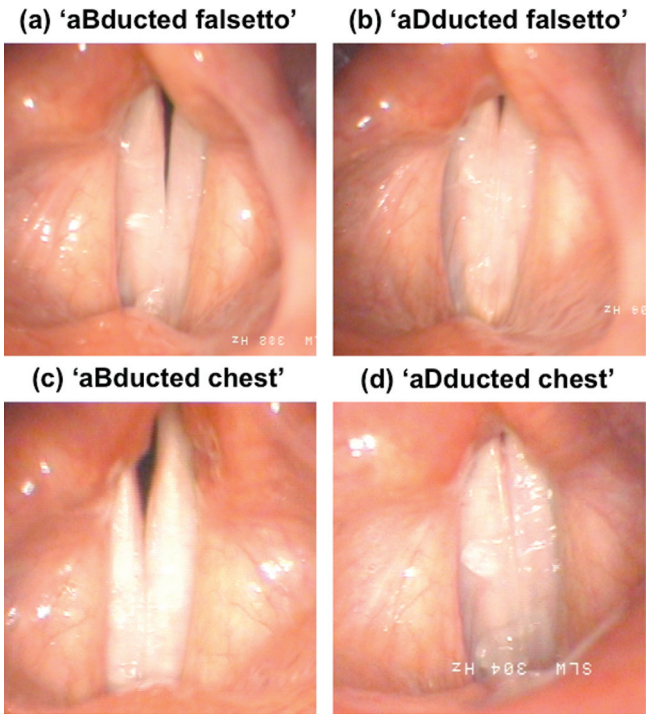


FIG. 1. (Color online) Typical laryngeal configurations for all attempted phonation types (female subject MM) as revealed by laryngeal videostroboscopy. (a) aBducted falsetto, (b) aDducted falsetto, (c) aBducted chest, and (d) aDducted chest. The images were taken at the moment of maximal vocal fold closure.

fundamental frequency. For each of the samples of stable phonation, the CQ_{VKG} was calculated from the three consecutive glottal cycles located at half the duration of the target note.

For both CQ_{VKG} and glottal area data, the difference among all phonation types was evaluated statistically by analysis of variance (ANOVA, software SIGMASTAT v.3, SPSS, USA). After ANOVA revealing statistically significant difference ($p < 0.05$), multiple comparisons were performed using the *t*-test with Bonferroni corrections (e.g. Abdi, 2007). A critical level of 0.01 was chosen for this purpose.

III. RESULTS

All the subjects had a greater PGC and thus a less adducted posterior glottis in the two aBducted (breathy) phonation types ("aBducted falsetto" and "aBducted chest") than in the two aDducted (non-breathy) phonation types ("aDducted falsetto" and "full chest"). These findings are demonstrated in the examples of laryngostroboscopic images which are provided in Fig. 1—notice the larger posterior glottal gap in the aBducted phonation types [Figs. 1(a) and 1(c)] than in the aDducted types [Figs. 1(b) and 1(d)].

For evaluating the differences between the chest and falsetto phonations, the VKG images were used. In Fig. 2 it can be seen that the duration of the glottal closure (investigated approximately in the middle of the membranous portion of the glottis, where the vibration amplitude was greatest) is considerably shorter in falsetto [Fig. 2(a)] than in chest register [Fig. 2(c)] for the more aBducted phonation, and the same is true when comparing the more aDducted phonations [Figs. 2(b) vs. 2(d)]. This visual impression was confirmed by the measurement of the CQ from the videokymographic images CQ_{VKG} (Fig. 3, see text further on).

The duration of glottal closure varied within both the chest and falsetto registers when changing the posterior adduction, as expected. The closed quotient CQ_{VKG} rose when changing from aBducted falsetto to aDducted falsetto [Figs. 2(a) vs. 2(b)], as well as when changing from aBducted chest to aDducted chest [Figs. 2(c) vs. 2(d)]. Interestingly, in five subjects (HM, JS, KP, ME, and WB) the CQ_{VKG} values reached similar or even larger values for the FaD than for the CaB [see Figs. 4 and 5(d) and the text further on].

Besides the CQ_{VKG} values, differences between the chest and falsetto registers were also observed in the VKG images; namely in the sharpness of the lateral peak, which is a visual feature related to the vertical phase difference in the vocal fold mucosal movement (Sundberg and Högset, 2001; Švec *et al*., 2007). Visually, the VKG images of chest phonations generally had sharper lateral peaks than those of falsetto phonation, suggesting larger vertical differences in chest than in falsetto register, but no quantitative data on these phenomena could be obtained for this study.

For subject HS, who had bowed vocal folds, no distinct differences between falsetto and chest phonations (FaB vs CaB, and FaD vs CaD) were perceivable. Subject MD had a limited falsetto function due to vocal-fold asymmetry and sulcus vocalis. Due to those reasons, falsetto phonations of subjects HS and MD have been excluded from the analysis.

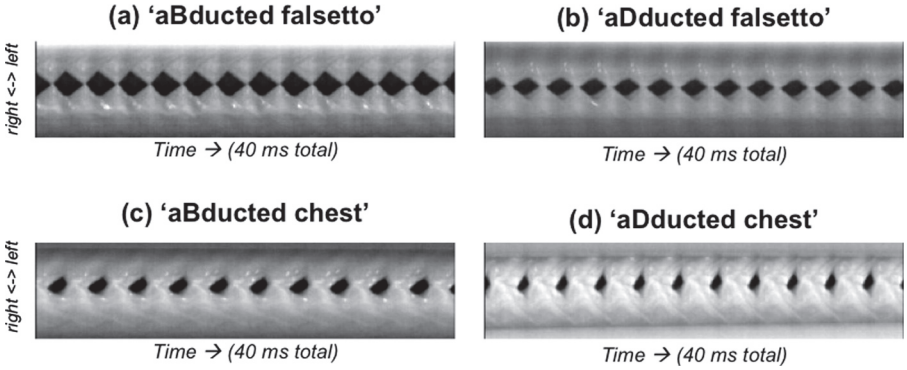


FIG. 2. Typical VGK images of all attempted phonation types for male subject QQ. (a) aBducted falsetto, (b) aDducted falsetto, (c) aBducted chest, and (d) aDducted chest. All the VKG images were recorded at the place of maximal vibration amplitude of the vocal folds (total time displayed: 40 ms).

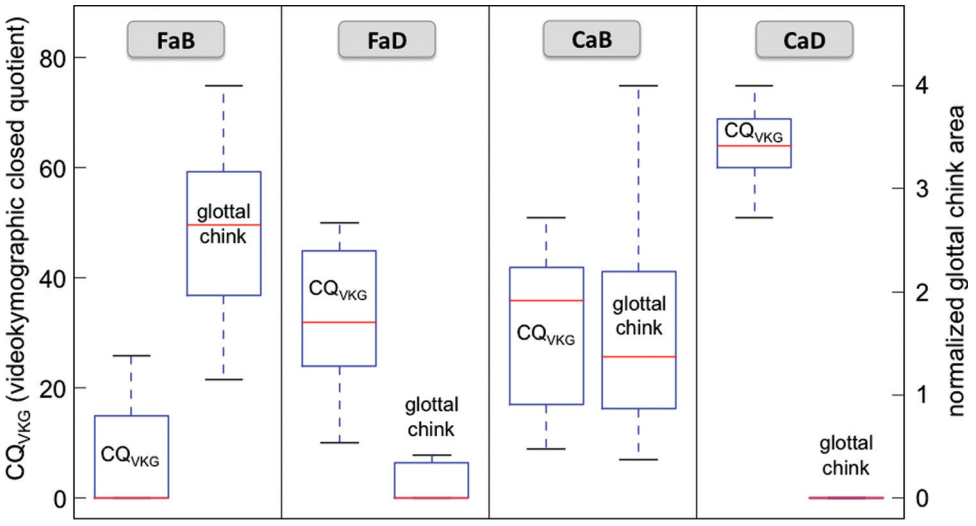


FIG. 3. (Color online) Box plot displays of analysis data. For each phonation type, CQ_{VKG} is displayed as the left column and glottal chink area is displayed as the right column.

Due to limited visibility, the glottal area could not be determined for the chest phonations of subject WB, and was thus also excluded from the analysis.

ANOVA revealed statistically significant differences ($p < 0.05$) for the data sets of both the CQ_{VKG} and glottal area. To find out which pairs of phonation types were significantly different from each other, multiple comparisons were performed using the *t*-test with Bonferroni corrections (Bt-test). The distribution of videokymographic closed quotient CQ_{VKG} and glottal chink area data for all subjects is shown in Fig. 3. The box plots illustrate that the aBducted phonation types (FaB and CaB) were generally produced with a larger PGC than their aDducted counterparts (FaD and CaD), as expected. The difference was statistically significant (Bt-test, $p < 0.001$). The chest phonations (CaB and CaD) had a significantly larger closed quotient than their respective falsetto counterparts (FaB and FaD) (Bt-test, $p < 0.001$).

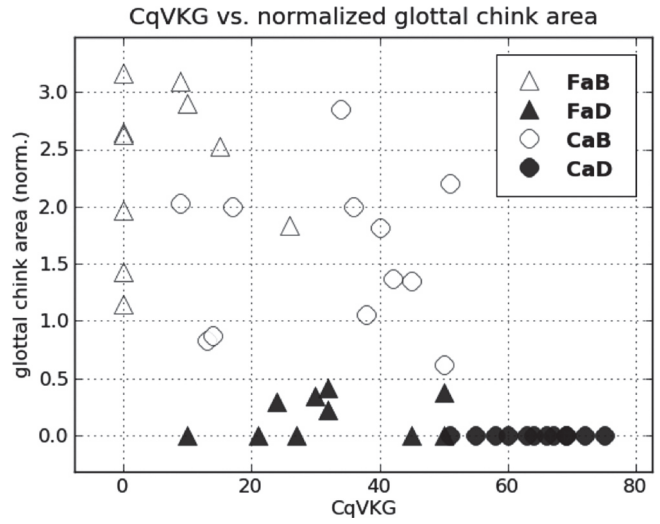


FIG. 4. Graphical illustration of the relation of CQ_{VKG} to glottal chink area for all subjects. Falsetto phonations are shown as triangles and chest phonations are shown as circles, respectively. Empty symbols represent aBducted phonations and filled symbols represent aDducted phonations, respectively. Note that the aBducted falsetto phonations are found in the upper left corner (signifying a small closed quotient and a large glottal chink area) and the aDducted chest phonations are located in the lower right corner (representing a fully adducted posterior glottis and a large closed quotient).

The comparison of videokymographic closed quotient CQ_{VKG} and glottal chink area data for each individual phonation type and each subject is shown in Fig. 4. Here, all the aDducted chest phonations are found in the lower right corner of the graph (having a high CQ_{VKG} and no PGC, i.e. fully adducted posterior glottis). On the other hand, the aBducted falsetto phonations are located in the upper left corner of the graph, since they were generally produced with a large posterior glottal gap and a small CQ_{VKG}. The aDducted falsetto phonations and the aBducted chest phonations occupy the central area of the graph. They can be distinguished from each other by the difference in PGC area (the aBducted chest phonations generally showing a larger PGC).

In Fig. 5, the transitions between the phonation types are shown for all subjects. The subplots [Figs. 5(a)–(f)] reveal that:

(a) To change from aBducted falsetto to aDducted falsetto, all subjects increased membranous as well as cartilaginous adduction (all arrows rightwards and downwards). The Bonferroni *t*-test (Bt-test) revealed that the differences between the CQ_{VKG} values and the chink areas were both statistically significant ($p < 0.001$).
(b) To change from aBducted falsetto to aBducted chest register, all subjects increased the membranous adduction (arrows rightwards). The difference showed to be statistically significant in the CQ_{VKG} values (Bt-test, $p < 0.001$). The posterior adduction was not decisive for this change (three subjects increased, seven subjects decreased posterior chink, and two remained about the same)—statistically, the difference in the chink areas did not reach the required significance level of 0.01 (Bt-test, $p = 0.012$).
(c) To change from aBducted falsetto to aDducted chest, all subjects increased membranous as well as cartilaginous adduction (all arrows rightwards and downwards). Both

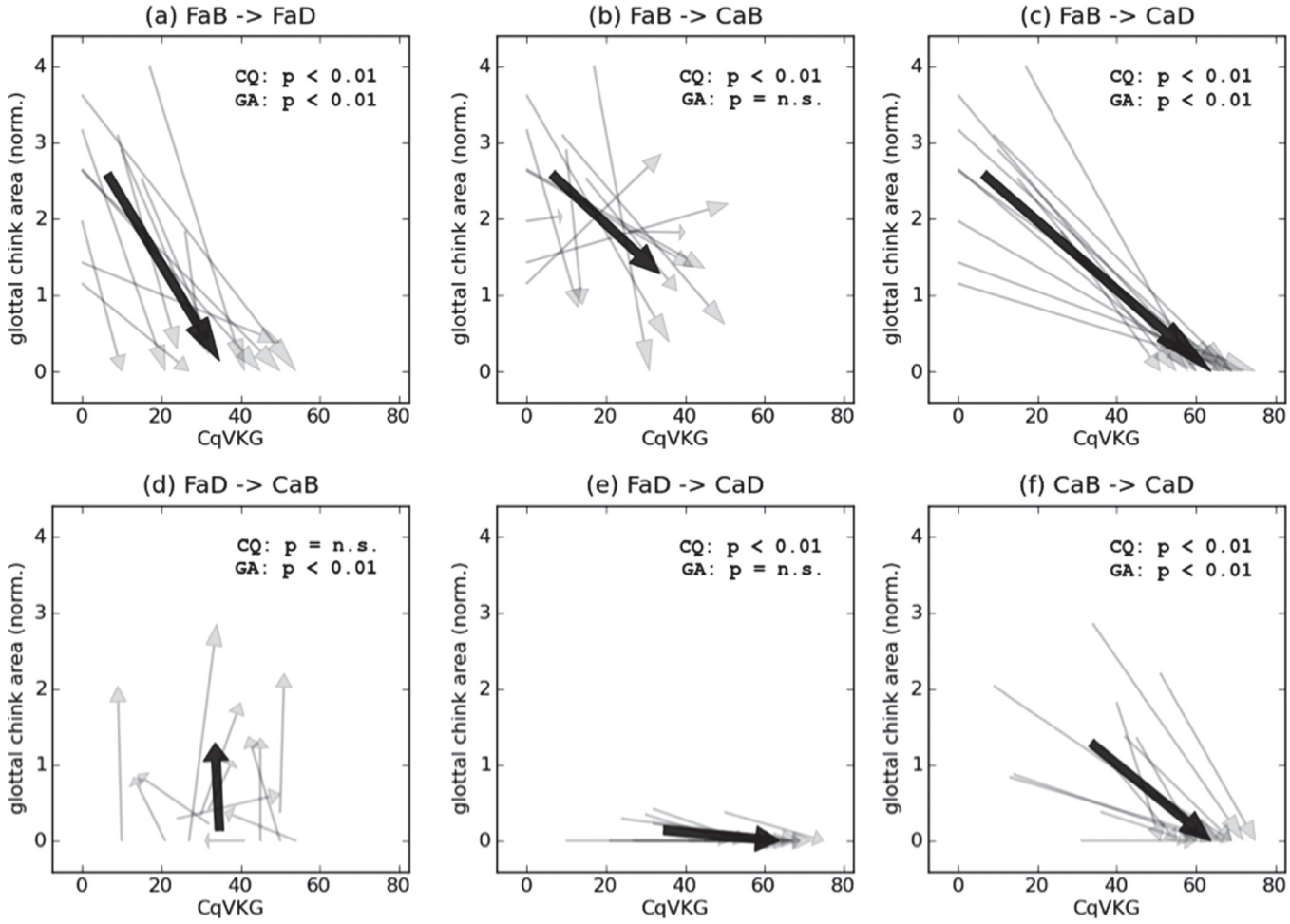


FIG. 5. Changes of laryngeal settings (represented by CQ_{VKG} and glottal chink area) for transitions between phonation types for all subjects: (a) aBducted falsetto to aDducted falsetto, (b) aBducted falsetto to aBducted chest, (c) aBducted falsetto to aDducted chest, (d) aDducted falsetto to aBducted chest, (e) aDducted falsetto to aDducted chest, (f) aBducted chest to aDducted chest. The large arrow in each sub-plot represents the general trend based on the average values of all subjects per phonation type. Since some phonations had to be excluded from the analysis (see text for details), some sub-plots contain fewer arrows than the number of subjects participating in the experiments. The *p*-values of the Bonferroni *t*-test are displayed in the upper right corner of each sub-plot. Abbreviations used: GA, glottal area; CQ_{VKG}, videokymographic closed quotient; n.s., not significant.

the CQ_{VKG} values and the chink areas showed statistically significant differences (Bt-test, $p < 0.001$).

(d) To change from aDducted falsetto to aBducted chest register, all subjects decreased posterior adduction (chink areas: Bt-test, $p < 0.001$). Membranous adduction (i.e., the CQ_{VKG}) was not decisive for this change: five decreased, four increased, and three remained about the same. Statistically, the CQ_{VKG} differences were not significant ($p > 0.05$).

(e) To change from aDducted falsetto to aDducted chest, all subjects increased the membranous adduction (all arrows pointing to the right—the CQ_{VKG} values were significantly different, Bt-test, $p < 0.001$). The cartilaginous adduction remained either maximal (five cases) or was further tightened (five cases). The chink area differences were not statistically significant (Bt-test, $p > 0.5$).

(f) To change from aBducted chest to aDducted chest, all subjects increased both membranous and cartilaginous adduction. Both the CQ_{VKG} values and the chink areas showed statistically significant differences (Bt-test, $p < 0.001$).

IV. DISCUSSION

The results showed distinct laryngeal adjustments for the four phonation types in most of the subjects. The most remarkable factor for distinguishing the phonation types FaB from FaD, and CaB from CaD (i.e., the "abducted" from the "adducted" phonations) was found to be the existence and area of the PGC. In the aDducted (FaD and CaD) phonation types the posterior glottis was either fully adducted or showed only a small cartilaginous PGC. On the other hand, in the aBducted (FaB and CaB) phonation types, the posterior glottis was always more open than in the aDducted phonation types [Figs. 5(a), 5(c), 5(d), 5(f)]. Inferential statistics revealed significant differences of glottal area between phonations that were targeted to be produced with different posterior glottal configurations (i.e., "aBducted" vs "aDducted"). No significant difference, however, was found in the data of glottal chink area for phonations produced with similar posterior glottal configurations (i.e., aBducted falsetto vs aBducted chest; and aDducted falsetto vs aDducted chest) [Figs. 5(b) and 5(e)].

For distinguishing the phonation types FaB from CaB, and FaD from CaD (i.e., the aBducted falsetto from the aBducted chest and the aDducted falsetto from the aDducted chest) videokymography was found useful. Specifically, the chest phonations exhibited a longer closed phase (i.e., larger CQ_{VKG}) than the respective falsetto phonations [recall Figs. 5(b) and 5(e)]. According to previous studies, this can be contributed to the activity of the vocalis muscle which thickens the body of the vocal fold (and therefore aDducts the membranous portion of the vocal folds) and slackens the vocal fold cover (Hirano, 1974; Titze, 2000), thus increasing the vertical phase difference in vocal fold vibration.

It has been generally accepted that the duration of the closed phase is greater in chest register than in falsetto (Choi *et al.*, 1993; Henrich *et al.*, 2005; Hirano, 1981; Roubeau, 2009; Salomão and Sundberg, 2008; Vilkman *et al.*, 1995). Our data, based on CQ_{VKG} calculations, indicate that this assumption is valid, but only for the cases when the posterior adduction in falsetto is the same or reduced compared to the chest register. In cases when the chest register is produced with an aBducted posterior glottis, the CQ_{VKG} values may sometimes reach equal or smaller values than in aDducted falsetto phonations. This trend was observed here in eight subjects: in five subjects the CQ_{VKG} decreased [five arrows pointing leftwards in Fig. 5(d)] and in three more subjects the CQ_{VKG} remained about the same [three arrows pointing straight upwards in Fig. 5(d)]. According to our knowledge, such a finding has not been reported before. It implies that the closed quotient (at the place of maximum vibration amplitude) should not be used as a sole indicator of voice registers in singing.

In order to interpret the findings, it is useful to consider the relationship between membranous and cartilaginous adduction. When the arytenoid cartilages (and particularly their vocal processes) are adducted, the area of the PGC is reduced. Simultaneously, the membranous portions of the vocal folds are adducted since the vocal folds are posteriorly attached to the arytenoid cartilages at the vocal processes. The membranous adduction is, however, also influenced by another adjustment of the vocal folds, which we call *membranous medialization through vocal fold bulging* (short: *membranous medialization*). This adjustment is caused by an active increase of the volume of the membranous portion of the vocal folds through the activity and bulging of the TA muscle. The membranous adduction thus can be targeted separately through the membranous medialization, without largely affecting the degree of cartilaginous adduction (Nasri *et al.*, 1994). Lindestad and Södersten (1988) called this adjustment as "centrally located medial compression of the vocal folds." They reported a greater compression of the vocal folds in baritone voice (chest register) than in counter-tenor voice (falsetto register), i.e., the vocal folds appeared to be more tightly in contact during baritone voice phonation. This "centrally located medial compression of the vocal folds" (controlled by the contraction of the TA muscle) is to be distinguished from the posterior "medial compression" (controlled by the LCA muscles) described by van den Berg (1960); the latter of which would be comparable to cartilaginous adduction as described in this paper.

Based on anatomical and physiological knowledge (Nasri *et al.*, 1994; Zemlin, 1997), it can be assumed that these (intrinsic laryngeal) muscles are involved in changing the position of the membranous and cartilaginous portions of the vocal folds:

(1) TA [thyroarytenoid (vocalis) muscles]: Thickening and bulging the membranous portion of the vocal folds;
(2) LCA (lateral cricoarytenoid muscles): aDducting the arytenoids, particularly the vocal processes, and consequently aDducting also the membranous portion of the vocal folds;
(3) IA (interarytenoid muscles): Known to pull the posterior side of the arytenoid cartilages together, aDducting the posterior cartilaginous glottis, possibly slightly aBducting the membranous portion of the vocal folds;
(4) PCA (posterior cricoarytenoid muscles): aBducting both the arytenoids and the membranous portion of the vocal folds.

Different singers can potentially engage these muscles at different degrees of contraction. Our results, however, do not allow specifying to which degree each of these muscles was involved in our experiments. The exercises and experimental setup used in this study aimed only at distinguishing membranous medialization (choice of register) from posterior adduction (degree/absence of breathiness), in order to find out whether they can be manipulated separately. The possibility of even finer control over the individual muscles would require designing much more complex singing exercises. These were beyond the scope of our study.

The complex relationship between cartilaginous and membranous adduction is well observable in Fig. 5: A pure increase of posterior adduction (without changing the register and thus not employing the membranous medialization) would result in an increase of membranous adduction leading to increased CQ_{VKG} [see Figs. 5(a) (FaB → FaD) and 5(f) (CaB → CaD)]. The effect of membranous medialization is revealed in Fig. 5(e) (FaD → CaD): The closed quotient changes without changing the posterior adduction (horizontal orientation of the arrows). When decreasing cartilaginous adduction and increasing membranous medialization (via the choice of register), the effects on membranous adduction would even out, as illustrated in Fig. 5(d) (FaD → CaB, arrows pointing mainly upwards indicate little change of the closed quotient but large change of posterior adduction). This was also reflected by our statistical data: For aDducted falsetto vs aBducted chest no significant differences were found for the closed quotients, but the PGCs were significantly different.

According to the analyzed data, the designed phonatory exercises allowed the subjects to separately manipulate the cartilaginous adduction (by changing the degree of posterior vocal fold adduction) and membranous medialization (by changing the register of phonation). These results suggest that the four targeted phonation types produced by the subjects were created by varying the degrees of cartilaginous adduction and membranous medialization:

(1) aBducted falsetto (FaB): Less cartilaginous adduction, less membranous medialization;

(2) aDducted falsetto (FaD): More cartilaginous adduction, less membranous medialization;
(3) aBducted chest (CaB): Less cartilaginous adduction, more membranous medialization;
(4) aDducted chest (CaD): More cartilaginous adduction, more membranous medialization.

Schematically, the adduction and glottal configuration of the four phonation types is displayed in Fig. 6.

The importance of vocal fold adduction in voice production has been mentioned in earlier research: Garcia (1847) observed that in singing, the arytenoids can be "vigorously pinched together," or they can be separated, causing the glottis to form an "isosceles triangle, the little side of which is formed between the arytenoids." Those two glottal configurations cause tones of "very pronounced brilliance" and "extremely dull notes," respectively. Rubin and Hirt (1960) identified three basic patterns of vibratory activity in the falsetto register: "Open-chink," in which the cordal margins of the vocal folds do not or only occasionally touch each other in vibration; "closed-chink," in which they do make contact; and "damping," in which pitch is raised by progressive approximation of corresponding segments of the vocal folds usually from posterior to anterior. In counter-tenor voice complete glottal closure and a mucosal wave were found in the "high intensity" phonations, while a less pronounced mucosal wave and incomplete glottal closure were seen in some counter-tenors when phonating at low intensities (Södersten and Lindestad, 1987).

A similar configuration shown for the aBducted chest phonation in Fig.6 has also been found in patients with a posterior glottal gap, who compensate for the glottal insufficiency by bulging the membranous portion of the vocal folds (diagnosed as hyper-functional breathy voice or muscle tension dysphonia) (Morrison *et al.*, 1983). According to our results, such a configuration can also be achieved voluntarily in singing (without pathologic implications) to achieve a specific sound quality. The degree of glottal closure is likely to have an effect on the quality of the formants measured in the acoustic signal radiated from the mouth. (Barney *et al.*, 2007; Hanson, 1997). This effect has not been investigated here, but may be an interesting subject for future research.

Verdolini *et al.* (1998) found that resonant voice is produced "with a barely adducted or barely abducted laryngeal configuration that was distinct from configurations for pressed and breathy (but not normal) voices." The abduction quotient (i.e., the relationship between the prephonatory glottal half width and the amplitude of vocal fold vibration) as an indicator of spectral slope (and thus register) is related to the degree of TA muscle contraction and inversely related to the degree of separation of the vocal processes (Titze, 2000). Švec *et al.* (2008) observed a small gap in the posterior glottis during phonation of an untrained female in head register, while no glottal gap was evident in the chest register, suggesting that muscles adducting the cartilaginous part of the glottis were slightly released when performing the transition from the chest to the head register.

Despite of these previous studies, to the best of our knowledge, this study is the first one that describes the role of different types of vocal fold adduction (cartilaginous vs membranous) as adjustable physiologic parameters in singing systematically. Based on our data and experience, we consider the individual control over cartilaginous vocal fold adduction vs membranous vocal fold medialization to be one of the key factors for an experienced singer to create different vocal timbres. It is conceivable that the lack of independent control over those two types of adduction may lead to problems in students who fail in singing specific styles.

Our study used both trained and non-trained singers as subjects. Trained singers are expected to have a finer control over minute laryngeal adjustments in singing than untrained singers and as such they are considered more reliable subjects for singing voice research. This study showed, however, that also four out of five non-trained singers were capable of separately adjusting the cartilaginous and membranous adduction using the exercises employed here. It may therefore be concluded that such adjustments are not

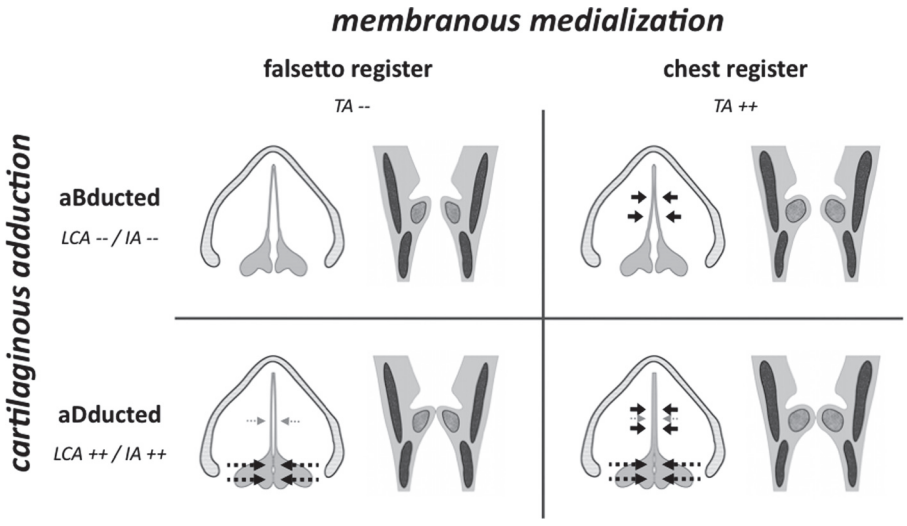


FIG. 6. Schematic illustration of the effect of *cartilaginous adduction* and *membranous medialization through vocal fold bulging* in singing. For each adduction type, two schematic graphs are shown: top view of vocal folds, arytenoids and thyroid cartilage (left); sagittal view of larynx with schematic drawings of thyroid cartilage, cricoid cartilage, and TA muscle (right). The arrows indicate the primary changes in the vocal fold position for each case.

limited to trained singers. However, a limiting factor is the ability to produce falsetto register. Two of the subjects had to be excluded from parts of the analysis due to this problem.

Further research is necessary to investigate the effect of different types of vocal fold adduction on glottal airflow and the spectrum of the voice source (and hence vocal timbre). In particular, it needs to be clarified how singers use this physiological parameter during performance for artistic expression in a stage situation; and whether certain voice types (lyrical vs dramatic 'Fach') use different base settings for vocal fold adduction.

The goal of this study was to explore the possibilities of a separate adjustment of the cartilaginous and membranous vocal fold adduction in a number of singers and non-singers. No attempt was made to investigate the underlying activity of intrinsic laryngeal muscles. In particular, one potential limitation of this study might be the fact that no distinction was made between activation of the LCA muscle (which rotates the arytenoid cartilages to bring the vocal processes toward the glottal midline) and the IA muscles (IA which approximate the arytenoid cartilages) (Zemlin, 1997). Moreover, the antagonistic action of the posterior cricoarytenoid muscle (which is actually an aBductor of vocal folds) is also likely to play a role in fine-tuning cartilaginous adduction in singing. In order to further understand the contribution of the individual muscles to the voice quality in singing, future electromyographic (EMG) studies are needed, and the gathered data should be related to computational models for vocal fold posturing (e.g., Titze and Hunter, 2007).

V. CONCLUSIONS

The obtained data supports the initial hypothesis that the singers and non-singers (of both sexes) can independently control the *cartilaginous adduction* and *membranous medialization* (by means of adduction of the arytenoid cartilages and the bulging of the vocal folds, respectively). To the best of our knowledge, such a finding has not been documented before. Independent control over cartilaginous adduction and membranous medialization is particularly important for the singing voice, helping the experienced singer to fine-tune the characteristics of the sound source. In this respect, the exercises described here can be useful for singers with limited flexibility who experience problems with producing different timbres. The gathered data also showed that the videokymographically derived closed quotient should not be considered to be the sole indicator of the voice register in singing, since it can in some subjects achieve larger values in "aDducted falsetto" than in "aBducted chest" phonations. This is attributed to the fact that membranous adduction (and thus the closed quotient) is influenced by both membranous medialization and cartilaginous adduction.

ACKNOWLEDGMENTS

In the Netherlands, the research was supported by the Technology Foundation, STW (Stichting Technische Wetenschappen) project GKG5973, Applied Science Division of NWO (Natuurwetenschappelijk Onderzoek), and the technology program of the Ministry of Economic Affairs, the Netherlands. In the Czech Republic, the work was supported by the Grant Agency of the Czech Republic, project GAČR 101/08/1155. We kindly thank Dr. D. Lazár from the Department of Biophysics, Palacký University Olomouc, for his help with the statistic analysis of the data. We thank the anonymous reviewers for their feedback and their helpful comments.

Abdi, H. (**2007**). "Bonferroni and Šidák corrections for multiple comparisons," In *Encyclopedia of Measurement and Statistics*, edited by Salkind, N. J. (Sage Publications: Thousand Oaks, CA). Available online at http://www.utdallas.edu/~herve/Abdi-Bonferroni2007-pretty.pdf (Last viewed December 22, 2010).

Baken, R., and Isshiki, N. (**1977**). "Arytenoid displacement by simulated intrinsic muscle contraction," Folia Phoniatr. **29**, 206–216.

Baken, R. J., and Orlikoff, R. F. (**2000**), *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Publishing, Thompson Learning, San Diego, CA), Chap. 10 , pp. 393–451

Barlow, C., and Howard, D. M. (**2005**). "Electrolaryngographically derived voice source changes of child and adolescent singers," Logoped. Phoniatr. Vocol. **30**(3,4), 147–157.

Barney, A., De Stefano, A., Henrich, N. (**2007**). "The effect of glottal opening on the acoustic response of the vocal tract," Acta. Acust. Acust. **93**, 1046–1056

Berg van den, Jw., Vennard, W., Burger, D., and Shervanian, C. C. (**1960**). Voice Production. The Vibrating Larynx, Instructional film. (University of Groningen, the Netherlands), p. 31.

Broad, D. (**1968**). "Kinematic considerations for evaluating laryngeal cartilage motions," Folia Phoniatr. **20**, 269–284.

Bunch, M., and Chapman, J. (**2000**). "Taxonomy of singers used as subjects in scientific research," J. Voice **14**(3), 363–369.

Choi, H. S., Berke, G. S., Ye, M., and Kreiman, J. (**1993**). "Function of the thyroarytenoid muscle in a canine laryngeal model," Ann. Otol. Rhinol. Laryngol. **102**(10), 769–776.

Echternach, M., Sundberg, J., Arndt, S., Breyer, T., Markl, M., Schumacher, M., and Richter, B. (**2008**). "Vocal tract and register changes analysed by real-time MRI in male professional singers – A pilot study," Logoped. Phoniatr. Vocol. **33**(2), 67–73.

Echternach, M., Sundberg, J., Arndt, S., Markl, M., Schumacher, M., and Richter, B. (**2010**). "Vocal tract in female registers – A dynamic real-time MRI study," J. Voice **24**(2), 133–139.

Fried, M. P., Meller, S. M., and A. Rinaldo, A. (**2009**). "Adult laryngeal anatomy," in *The Larynx*, edited by Fried, M. P., and Ferlito, A. (Plural Publishing, San Diego, CA), p. 85–99.

Garcia, M. (**1847**). "Traité complet de l'art du chant (A complete treatise on the art of singing)" (Minkoff Éditeur), as cited by Stark, J. (**2003**). *Bel Canto. A History of Vocal Pedagogy*. (University of Toronto Press, Toronto, CA), Chap. 1, pp. 3–32.

Gullaer, I., Walker, R., Badin, P., and Lamalle, L. (**2006**). "Image, imagination, and reality: On effectiveness of introductory work with vocalists," Logoped. Phoniatr. Vocol. **31**(2), 89–96.

Hanson, H. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J. Acoust. Soc. Am. **115**(3), 1321–1332.

Henrich, N., d'Alessandro, C., Doval, B., and Castellengo, M. (**2005**). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency," J. Acoust. Soc. Am. **117**(3), 1417–1430.

Henrich, N., Kiek, M., Smith, J., and Wolfe, J. (**2007**). "Resonance strategies used in Bulgarian women's singing style: A pilot study," Logoped. Phoniatr. Vocol. **32**(4), 171–177.

Herbst, C. T., and Ternström, S. (**2006**). "A comparison of different methods to measure the EGG contact quotient," Logoped. Phoniatr. Vocol. **31**(3), 126–138.

Herbst, C. T., Ternström, S., and Švec, J. G. (**2009**). "Investigation of four distinct glottal configurations in classical singing – A pilot study," J. Acoust. Soc. Am. **125**(3), EL104–EL109. Acoustic and laryngoscopic samples of those four phonation types are freely available at http://dx.doi.org/10.1121/1.3057860 (Last viewed March 8, 2010).

Hirano, M. (**1974**). "Morphological structure of the vocal cord as a vibrator and its variations," Folia Phoniatr. **26**, 89–94.
Hirano, M. (**1981**), *Clinical Examination of Voice* (Springer-Verlag, New York), p. 100.
Hunter, E., Titze, I. R., and Alipour, F. (**2004**). "A three-dimensional model of vocal fold abduction/adduction," J. Acoust. Soc. Am. **115**(4), 1747–1759.
Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**(4 Pt 1), 2434–2439.
Ladefoged, P. (**1975**). *A Course in Phonetics* (Harcourt, Orlando, FL), Chap. 6 , pp. 113–135.
Letson, J. A., Jr., and Tatchell, R. (**1997**) "Arytenoid movement," in *Professional Voice: The Science and Art of Clinical Care*, edited by Sataloff, R. T. (Singular Publishing Group, San Diego, CA), p. 131–145.
Lindestad, P. A., Södersten, M. (**1988**) "Laryngeal and pharyngeal behavior incountertenor and baritone singing – A videofiberscopic study," J. Voice **2**(2), 132–139.
Miller, D. G., Švec, J. G, and. Schutte, H. K. (**2002**). "Measurement of characteristic leap interval between chest and falsetto registers," J. Voice **16**(1), 8–19.
Moore, P., and von Leden, H. (**1958**). "Dynamic variations of the vibratory pattern in the normal larynx," Folia Phoniatr. **10**(4), 205–238.
Morrison, M. D., Rammage, L. A., Belisle, G. M., Pullan, C. B., and Nichol, H. (**1983**). "Muscular tension dysphonia," J. Otolaryngol. **12**, 302–306.
Nasri, S., Sercarz, J. A., Azizzadeh, B., Kreiman, J., and Berke, G. S. (**1994**). "Measurement of adductory force of individual laryngeal muscles in an in vivo canine model," Laryngoscope **104**, 1213–1218.
Qiu, Q., and Schutte, H. K. (**2006**). "A new generation videokymography for routine clinical vocal-fold examination," Laryngoscope **116**(10), 1824–1828.
Qiu, Q., Schutte, H. K., Gu, L., and Yu, Q. (**2003**). "An automatic method to quantify the vibration properties of human vocal folds via videokymography," Folia Phoniatr Logoped. **55**(3), 128–136.
Roubeau, B., Henrich, N., and Castellengo, M. (**2009**). "Laryngeal vibratory mechanisms: The notion of vocal register revisited," J. Voice **23**, 425–438.
Rubin, H. J., and Hirt, C. C. (**1960**). "The falsetto. A high-speech cinematographic study," Laryngoscope **70**, 1305–1324.
Salomão, G. L., and Sundberg, J. (**2008**). " Relation between perceived voice register and flow glottogram parameters in males," J. Acoust. Soc. Am. **124**(1), 546–551.
Salomão, G. L., and Sundberg, J. (**2009**). "What do male singers mean by modal and falsetto register? An investigation of the glottal voice source," Logoped. Phoniatr. Vocol **34**(2), 73–83.
Sataloff, R. T., (**1997**). "Clinical anatomy and physiology of the voice," in *Professional Voice: The Science and Art of Clinical Care* edited by Sataloff, R.T. (Singular Publishing Group, San Diego, CA), pp. 111–130.
Schutte, H. K., Miller, D. G., and Duijnstee, M. (**2005**). "Resonance strategies revealed in recorded tenor high notes," Folia Phoniatr. Logop. **57**(5,6), 292–307.
Södersten, M., and Lindestad, P. A. (**1987**). "Vocal fold vibrations in countertenor singing." Acta Phon Lat. **9** (suppl.), 19–22.
Södersten, M., Hertegard, S., and Hammarberg, B. (**1995**). "Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women," J. Voice **9**(2), 182–197.
Story, B. H. (**2004**). "Vowel acoustics for speaking and singing," Acta. Acust. Acust. **90**(4), 629–640.
Sundberg, J. (**1972**). "An articulatory interpretation of the singing formant," Speech Transm. Lab. Q. Prog. Status Rep. **1**, 45–53.
Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant'," J. Acoust. Soc. Am. **55**(4), 838–844.
Sundberg, J., and Högset, C. (**2001**). "Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones," Logoped. Phoniatr. Vocol. **26**, 26–36.
Sundberg, J., and Skoog, J. (**1997**). "Dependence of jaw opening on pitch and vowel in singers," J. Voice **11**(3), 301–306.
Sundberg, J., Fahlstedt, E., and Morell, A. (**2005**). "Effects on the glottal voice source of vocal loudness variation in untrained female and male voices," J. Acoust. Soc. Am. **117**, 879–885.
Švec, J. G., Šram, F., and Schutte, H. K. (**2007**). "Videokymography in voice disorders: What to look for?," Ann. Otol. Rhinol. Laryngol. **116**(3), 172–180.
Švec, J. G., Sundberg, J., and Hertegard, S. (**2008**). "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," J. Acoust. Soc. Am. **123**(1), 347–353.
Timcke, R., von Leden, H., and Moore, P. (**1958**). "Laryngeal vibrations: Measurements of the glottic wave: Part I. The normal vibratory cycle," Arch. Otolaryngol. **68**, 1–19.
Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**(4), 1699–1707.
Titze, I. R. (**2000**), *Principles of Voice Production* (National Center for Voice and Speech, Denver, CO), p. 409.
Titze, I. R., and Hunter, E. J. (**2007**). "A two-dimensional biomechanical model of vocal fold posturing," J. Acoust. Soc. Am. **121**, 2254–2260
Verdolini, K., Druker, D., Palmer, P., and Samawi, H. (**1998**). "Laryngeal adduction in resonant voice," J. Voice **12**(3), 315–327.
Vilkman, E., Alku, P., and Laukkanen, A. M. (**1995**). "Vocal-fold collision mass as a differentiator between registers in the low-pitch range," J. Voice **9**, 66–73.
Wendler, J. (**2008**). "Singing and science," Folia Phoniatr. Logoped. **60**(6), 279–287.
Zemlin, W. R. (**1997**). *Speech and Hearing Science – Anatomy and Physiology* (Allyn and Bacon, Boston), Chap. 3 , pp. 100–196.

Supplement D: Python source code for generating EGG wavegrams

This section describes the content of two files: (a) the file wavegram.py, containing all the actual code for wavegram generation; and (b) the file test.py, which calls wavegram.py and contains all the matplotlib code to actually create a graph.

test.py

```
import wavegram
import matplotlib.pyplot as plt
import numpy

######################################################################
# open a two-channel file (right channel is EGG signal) and display
# both DEGG wavegram and raw EGG signal
######################################################################

# stereo WAVE file (left channel: acoustic signal; right channel: EGG signal)
fileName = 'tm_mdv_04_01_short.wav'

# change these values according to the analyzed EGG signal, in order to avoid
# clipping (in which case the colours will be either pure red or pure blue...),
# or use zero for both parameters to enable automatic scaling
eggDbMax = -22 # dB; represents pure red in the colour coding of the wavegram
eggDbMin = -24 # dB; represents pure blue in the colour coding of the wavegram

graphWidthInches = 10
graphHeightInches = 10
dpi = 72

######################################################################

# start processing
print "WAVEGRAM for file", fileName
data = wavegram.signal()
data.wrapWaveFile(fileName)

tStart = 0
tEnd = data.getDuration()

plt.clf()
fig = plt.figure(figsize=(graphWidthInches,graphHeightInches), dpi = dpi)

eggData = data.getRawData(1)
dataT = numpy.zeros(len(eggData))
fs = data.getSamplingFrequency()
for i in range(len(eggData)):
    dataT[i] = i / float(fs)
ax1 = plt.subplot(411)
ax1.plot(dataT, eggData)
ax1.grid(True)
ax1.set_xlabel('Time [sec]')
#ax1.set_title('EGG waveform')
ax1.set_ylabel('EGG waveform')
ax1.set_title(fileName)

waveGramData1 = data.createWavegram(tStart = tStart, tEnd = tEnd, \
    derivative = True, cycleDetectionTolerance = 0.005, \
    eggDbMax = eggDbMax, eggDbMin = eggDbMin)
ax2 = plt.subplot(412, sharex=ax1)
ax2.imshow(waveGramData1, interpolation='nearest', aspect='auto', \
    origin='lower', extent=[tStart, tEnd, 0, 100])
ax2.grid(True)
ax2.set_xlabel('Time [sec]')
ax2.set_ylabel('DEGG wavegram')

arrF0 = data.metaData.arrF0
```

```
ax3 = plt.subplot(413, sharex=ax1)
ax3.plot(arrF0['t'], arrF0['freq'], linewidth = 3)
#ax3.set_title('Fundamental frequency')
ax3.set_ylabel('F0 [Hz]')
ax3.set_xlabel('Time [sec]')
ax3.grid(True)
if tEnd == -1:
    tEnd = arrF0['t'][-1]
ax3.set_xlim(tStart, tEnd)

ax4 = plt.subplot(414, sharex=ax1)
dbDataX, dbDataY = data.calculateRmsVector(0.1, 0.05, 1, True)
ax4.plot(dbDataX, dbDataY, linewidth = 3)
#ax4.set_title('Amplitude')
ax4.set_ylabel('EGG Amplitude [dB rel.]')
ax4.set_xlabel('Time [sec]')
ax4.grid(True)
ax4.set_xlim(tStart, tEnd)

plt.subplots_adjust(hspace=0.5)
plt.show()
#fileNameOnly = '.'.join(fileName.split('.')[:-1])
#plt.savefig(fileNameOnly + '_wavegram.png')

print "done."
```

wavegram.py

```
import numpy
import math
import wave
import pickle
import matplotlib.pyplot as plt
import dspUtil

#http://www.ar.media.kyoto-u.ac.jp/members/david/softwares/audiolab/
import audiolab

##########################################################################
# here be helper functions
##########################################################################

def writeWaveFile(data, fileName, SRate = 44100.0, normalize = False):
    """
    write an array of floats (floatSignal) to a 16 bit wave file (fileName)
    using a specified sampling rate (44100). if the parameter normalize is set
    to True, the signal will be normalized to the maximally possible value
    (i.e. 1). if no normalization is performed, and if the input signal has a
    maximum absolut ampitude greater than 1 (i.e. if the output would be
    clipped), the function throws an error. The 'data' parameter of this
    function can be either an array of floats, or a list of arrays of floats.
    """

    isList = type(data).__name__ == 'list'
    if not isList:
        data = [ data ]

    # determine the file length and maximum amplitude
    numSamp = 0
    maxAmp = 0
    for floatSignal in data:
        numSamp += len(floatSignal)
        if floatSignal.max() > maxAmp:
            maxAmp = floatSignal.max()
        if math.fabs(floatSignal.min()) > maxAmp:
            maxAmp = math.fabs(floatSignal.min())

    # error checking
    if normalize == False:
        if maxAmp > 1:
            msg = "max. amplitude of given array is " + str(maxAmp) \
                + " - this would result in a clipped wave file signal"
            raise Exception(msg)
```

```
    # create the wave file and its header
    audioFile = wave.open (fileName, "wb")
    audioFile.setparams((1, 2, int(SRate), int(numSamp), "NONE", \
        "not compressed"))

    # write the data
    maxint = 2**15 - 1
    for floatSignal in data:
        if normalize:
            floatSignal = floatSignal / maxAmp # normalize here
        intWave = numpy.array(floatSignal * maxint, dtype = numpy.int16)
        audioFile.writeframes(intWave.tostring())

    # close the wave file
    audioFile.close()

########################################################################

def readWaveFile(fileName):
    w = wave.open(fileName, 'rb')
    data = ([w.getparams(), w.readframes(w.getnframes())])
    w.close()
    return data

########################################################################

def logger(msg):
    print "\t\tWAVEGRAM: " + msg

########################################################################

def interpolateLinear(
        y1, #
        y2, #
        x # weighting [0..1]. 0 would be 100 % y1, 1 would be 100 % y2
):
    """ simple linear interpolation between two variables """
    return y1 * (1.0 - x) + y2 * x

########################################################################

def interpolateParabolic(
        alpha,
        beta,
        gamma,
        x # relative position of read offset [-1..1]
):
    """ parabolic interpolation between three equally spaced values """
    if (x == 0): return beta

    #we want all numbers above zero ...
    offset = alpha;
    if (beta < offset): offset = beta
    if (gamma < offset): offset = gamma
    offset = math.fabs(offset) + 1

    alpha += offset;
    beta += offset;
    gamma += offset;

    a = b = c = 0;
    a = (alpha - 2.0 * beta + gamma) / 2.0
    if (a == 0):
        if (x > 1):
            return interpolateLinear(beta, gamma, x) - offset
        else:
            return interpolateLinear(alpha, beta, x + 1) - offset
    else:
        c = (alpha - gamma) / (4.0 * alpha)
        b = beta - a * c * c
        return (a * (x - c) * (x - c) + b) - offset

########################################################################

def findArrayMaximum(
        data,
        offsetLeft = 0,
        offsetRight = -1 # if -1, the array size will be used
):
```

```
    objType = type(data).__name__.strip()
    if objType <> "ndarray":
        raise Exception('data argument is no instance of numpy.array')
    size = len(data)
    if (size < 1):
        raise Exception('data array is empty')
    xOfMax = -1
    valMax = min(data)
    if offsetRight == -1:
        offsetRight = size
    for i in range(int(offsetLeft + 1), int(offsetRight - 1)):
        if data[i] >= data[i-1] and data[i] >= data[i + 1]:
            if data[i] > valMax:
                valMax = data[i]
                xOfMax = i
    if xOfMax > 0 and xOfMax < size - 1:
        # use parabolic interpolation to increase accuracy of result
        alpha = data[xOfMax - 1]
        beta = data[xOfMax]
        gamma = data[xOfMax + 1]
        xTmp = (alpha - gamma) / (alpha - beta * 2 + gamma) / 2.0
        xOfMax = xTmp + xOfMax
        valMax = interpolateParabolic(alpha, beta, gamma, xTmp)
    return [xOfMax, valMax]

######################################################################

def corr(
        signal1, # input array
        signal2, # the other input array
        alignSize = True, # if true, the shorter array is scaled to the longer one
        zeroPaddingFactor = 2.0 # zero padding
):
    MAX_FFT_LENGTH = 2**20

    signalSize1 = len(signal1)
    signalSize2 = len(signal2)
    if alignSize == False:
        if signalSize1 <> signalSize2:
            raise Exception("the size of the two input signals does not match")

    signalSize = signalSize1
    if signalSize2 > signalSize: signalSize = signalSize2
    if signalSize <= 0 or signalSize >= MAX_FFT_LENGTH:
        raise Exception("the size of the input signals (" + str(signalSize) \
            + ") is out of range.")

    # make the actual window size a power of two
    targetWindowSize = int(signalSize * float(zeroPaddingFactor))
    realWindowSize = 1;
    while realWindowSize < targetWindowSize:
         realWindowSize *= 2

    if realWindowSize > MAX_FFT_LENGTH or realWindowSize < 1:
        raise Exception("array size (" + str(realWindowSize) \
            + ") is not allowed (range = 1 - " + str(MAX_FFT_LENGTH) + ").")

    arrIn1 = numpy.zeros(realWindowSize)
    arrIn2 = numpy.zeros(realWindowSize)
    arrIn3 = numpy.zeros(realWindowSize / 2)

    # copy the input data to the object's input data array
    # since the target array is longer, it is automatically zero-padded
    for k in range(signalSize1):
        arrIn1[k] = signal1[k]
    for k in range(signalSize2):
        arrIn2[k] = signal2[k]

    # calculate the FFTs
    arrFft1 = numpy.fft.rfft(arrIn1)
    arrFft2 = numpy.fft.rfft(arrIn2)

    # multiply result 1 with complex conjugate of result 2 and store it.
    for k in range(realWindowSize / 2):
        arrIn3[k] = arrFft1[k].conjugate() * arrFft2[k]

    # do the reverse fftp
    r = numpy.fft.irfft(arrIn3) * 2.0
```

```
    # favour smaller lags (avoid period doubling/tripling ...)
    for i in range(len(r)):
        r[i] *= 1.0 - i * 0.1 / len(r)

    # find the maximum in the lag function
    xOfMax, valMax = findArrayMaximum(r, signalSize * 0.25, signalSize - 1)

    return xOfMax * -1.0, r

######################################################################

def rmsToDb(
        rmsValue, # the RMS value that should be converted to a dB value
        dbBase = 0, # the dB base value
        rmsBase = 1.0 # the RMS base value
    ):
    """ performs a RMS to dB conversion """
    if rmsValue <= 0:
        raise Exception("RMS value must not be zero or below")
    return dbBase + (20.0 * math.log10(rmsValue / float(rmsBase)))

######################################################################

def getIdealEggWaveform(size):
    tmp   =   [   0.016129032,   0.118951613,   0.231854839,   0.336693548,   0.435483871,   0.524193548,
0.60483871,   0.675403226,   0.737903226,   0.790322581,   0.842741935,   0.885080645,   0.927419355,
0.953629032,  0.967741935,  0.977822581,  0.987903226,  0.991935484,  0.997983871,  1,  1,  0.997983871,
0.995967742,   0.989919355,   0.985887097,   0.977822581,   0.965725806,   0.955645161,   0.945564516,
0.933467742,  0.921370968,  0.909274194,  0.89516129,  0.875,  0.856854839,  0.838709677,  0.818548387,
0.798387097,   0.778225806,   0.756048387,   0.731854839,   0.705645161,   0.675403226,   0.64516129,
0.608870968,   0.570564516,   0.528225806,   0.483870968,   0.433467742,   0.370967742,   0.304435484,
0.213709677,   0.092741935,   0.008064516,   0.008064516,   0.008064516,   0.006048387,   0.006048387,
0.006048387,   0.006048387,   0.004032258,   0.004032258,   0.004032258,   0.004032258,   0.004032258,
0.004032258,   0.004032258,   0.004032258,   0.004032258,   0.004032258,   0.004032258,   0.002016129,
0.002016129,   0.002016129,   0.002016129,   0.002016129,   0.002016129,   0.002016129,   0.002016129,
0.002016129,   0.002016129,   0.002016129,   0.002016129,   0.002016129,   0.002016129,   0.002016129,
0.002016129, 0.002016129, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.002016129, 0.002016129 ]
    data = wavegramArray(tmp)
    data.stretch(size)
    return data

######################################################################



######################################################################
# array helper class
######################################################################

INTERPOLATE_NO = 0
INTERPOLATE_LINEAR = 1
INTERPOLATE_PARABOLIC = 2
INTERPOLATE_SINC = 3

class wavegramArray:
    """ This class wraps a numpy array. This design is necessary for two
        reasons: To make sure that array data can be passed as both a list or
        a numpy.array (and nothing else); and to implement a means for
        parabolic and sinc interpolation (the latter being needed for
        stretching arrays).
        TODO: change into a class derived from numpy.array
    """

    def __init__(self, data):
        self.init(data)

    # -------------------------------------------------------------------------- #

    def getSize(self):
        return self.size

    # -------------------------------------------------------------------------- #

    def init(self, data):
        objType = type(data).__name__.strip()
        if objType == "ndarray":
            self.data = data
            self.size = len(data)
        elif objType == "list":
```

```
            self.data = numpy.array(data)
            self.size = len(data)
        else:
            raise Exception('data argument is no instance of numpy.array')

    # ------------------------------------------------------------------- #

    def get(self, x, method = INTERPOLATE_LINEAR, sincWidth = 30):
        if (x > self.size and x < self.size + 1): x = self.size
        if (x < 0 or x > self.size):
            raise Exception('x (' + str(x) + ') is out of bounds')

        if (x >= self.size - 1):
            return self.data[self.size - 1]

        left = int(x) # force to int
        right = left + 1
        pos = float(x) - left # force to floating point
        if (pos == 0.0):
            return self.data[left]

        if method == INTERPOLATE_NO:
            # no interpolation: index is truncated to integral number
            return self.data[left]

        if method == INTERPOLATE_LINEAR:
            # simple linear interpolation
            return self.data[left] * (1.0 - pos) + self.data[right] * pos

        if method == INTERPOLATE_PARABOLIC:
            # parabolic interpolations
            left_val = 0
            right_val = 0
            if (left > 0):
                # need tocalculate the left val
                left_val = interpolateParabolic(self.data[left - 1], \
                    self.data[left], self.data[right], pos)
            else:
                left_val = interpolateLinear(self.data[0], self.data[1], pos)
            if (right < self.size - 1):
                # calculate the right val
                right_val = interpolateParabolic(self.data[left], \
                    self.data[right], self.data[right + 1], pos - 1)
            else:
                right_val = self.data[self.size - 2] * (1.0 - pos) \
                    + self.data[self.size - 1] * pos;
            return left_val * (1.0 - pos) + right_val * pos;

        if method == INTERPOLATE_SINC:
            # sinc interpolation
            tmp = 0
            for i in range(sincWidth):
                if (left - i >= 0):
                    # left-hand side
                    tmp += self.data[left - i] * numpy.sinc(pos + i)
                if (right + i < self.size):
                    # right-hand side
                    tmp += self.data[right + i] * numpy.sinc(pos - (1.0 + i))
            return tmp

        raise Exception("unknown interpolation method")

    # ------------------------------------------------------------------- #

    def stretch(self, newSize, method = INTERPOLATE_LINEAR, sincWidth = 30):
        newSize = int(newSize)
        dataNew = numpy.zeros(newSize)
        for i in range(newSize):
            idx = float(i) * (self.size + 1) / newSize
            dataNew[i] = self.get(idx, method, sincWidth)
        self.data = dataNew
        self.size = len(dataNew)



######################################################################
# signal and meta data
######################################################################
```

```
META_DATA_SUFFIX = '.meta'

class signalMetaData:

    def __init__(self):
        self.version = 1
        self.numFrames = 0
        self.fs = 0
        self.duration = 0
        self.audioChannel = 0
        self.eggChannel = 1
        self.f0ok = False
        self.F0progressPeriods = 0
        self.analysisTimeStep = 0
        self.arrF0 = { 't':[], 'freq':[] }
        self.arrSpl = []
        self.channels = 0
        self.encoding = ''
        self.glottalCyclesOk = False
        self.arrGlottalCycles = { 'offset':[], 'size':[] }
        pass

######################################################################

class signal:

    # TODO: I don't like the way the original WAVE data is currently stored.
    #       change accordingly.

    def __init__(self, data = [], fs = 44100):
        self.init(data, fs)

    # -------------------------------------------------------------------- #

    def init(self, data, fs):
        self.metaData = signalMetaData()
        # make sure data is of right type
        objType = type(data).__name__.strip()
        if objType == "ndarray":
            self.data = data # the data array is a numpy.array object
            self.metaData.numFrames = len(data)
        elif objType == "list":
            self.data = numpy.array(data) # the data array is a numpy.array object
            self.metaData.numFrames = len(data)
        else:
            raise Exception('data argument is no instance of numpy.array and no list')
        self.metaData.fs = fs
        self.metaData.duration = float(self.metaData.numFrames) / fs
        self.fileName = ''

    # -------------------------------------------------------------------- #

    def getDuration(self):
        return self.metaData.duration

    # -------------------------------------------------------------------- #

    def getRawData(self, channel):
        data = numpy.zeros(self.metaData.numFrames)
        for i in range(self.metaData.numFrames):
            data[i] = self.data[i][channel]
        return data

    # -------------------------------------------------------------------- #

    def getSamplingFrequency(self):
        return self.metaData.fs

    # -------------------------------------------------------------------- #

    def wrapWaveFile(self, fileName, verbose = False, fMin = 50,
            fMax = 3000, # don't look for F0 values above this value [Hz]
            voicingThreshold = 0.3, # autocorrelation-coefficient as criterion
                                    # for F0 detection; works just as in Praat
            targetChannel = 1, # F0 detection based on channel N [0..numChannels-1]
            f0DetectionApplyWindow = False, # whether or not to window the data
            glottalCycleDetectionTolerance = 1 # ms - see comment in
                                               # self.detectGlottalCycles(...)
    ):
```

```
        if verbose: logger('wrapping file ' + fileName)
        self.fileName = fileName
        f = audiolab.Sndfile(fileName, 'r')
        try:
            # analysis data (such as F0 information) is stored in a meta data
            # file together with the analyzed WAVE file. try to locate it, in
            # order to avoid re-calculating the analysis data.
            self.loadMetaData()
            tmp = signalMetaData()
            msg = 'Unable to restore meta data for file ' + fileName + ': '
            if self.metaData.version <> tmp.version:
                raise Exception(msg + 'meta data file version is outdated')
            if self.metaData.fs <> f.samplerate:
                raise Exception(msg + 'sample rates do not match (changed?)')
            if self.metaData.channels <> f.channels:
                raise Exception(msg + 'numChannels do not match (changed?)')
            if self.metaData.encoding <> f.encoding:
                raise Exception(msg + 'encoding does not match (changed?)')
            if self.metaData.numFrames <> f.nframes:
                raise Exception(msg + 'numFrames do not match (changed?)')
        except Exception as e:
            logger('WARNING: unable to load meta data: ' + e.__str__())
            self.metaData = signalMetaData()
            self.metaData.fs = f.samplerate
            self.metaData.channels = f.channels
            self.metaData.encoding = f.encoding
            self.metaData.numFrames = f.nframes
            self.saveMetaData()
        self.metaData.duration = float(self.metaData.numFrames) / self.metaData.fs
        self.data = f.read_frames(self.metaData.numFrames)
        f.close()
        if self.metaData.channels == 1:
            # in case of mono file, wrap data
            # that's ugly. change in new release!
            self.metaData.eggChannel = 0
            tmp = numpy.ones((len(self.data), 1))
            for i in range(len(self.data)):
                tmp[i][0] = self.data[i]
            self.data = tmp
        if self.metaData.f0ok == False:
            self.calculateF0(Fmin = fMin, Fmax = fMax, verbose=verbose, \
                voicingThreshold = voicingThreshold, \
                applyWindow = f0DetectionApplyWindow, \
                targetChannel = targetChannel)
        if self.metaData.glottalCyclesOk == False:
            self.detectGlottalCycles(verbose=verbose, \
                tolerance = glottalCycleDetectionTolerance)

    # ------------------------------------------------------------------------ #

    def calculateF0(self,
        Fmin = 50,
        Fmax = 3000,
        numPeriods = 5.0,
        progressPeriods = 1,
        targetChannel = -1, # if -1, the EGG channel is used
        verbose = False,
        voicingThreshold = 0.3,
        applyWindow = False
    ):
        if verbose: logger('calculating F0 for file ' + self.fileName)
        if targetChannel == -1: targetChannel = self.metaData.eggChannel
        if targetChannel >= self.metaData.channels: targetChannel = 0
        readSize = int(numPeriods * float(self.metaData.fs) / Fmin)
        offset = 0
        numCycles = 0
        self.metaData.arrF0 = { 't':[], 'freq':[] }
        self.metaData.F0progressPeriods = progressPeriods
        while (offset < self.metaData.numFrames):
            dataTmp = numpy.zeros(readSize)
            if self.metaData.channels == 1:
                for i in range(readSize):
                    idx = i + offset - (readSize / 2)
                    if idx >= 0 and idx < self.metaData.numFrames:
                        dataTmp[i] = self.data[idx]
            else:
                for i in range(readSize):
                    idx = i + offset - (readSize / 2)
                    if idx < self.metaData.numFrames:
```

```
                dataTmp[i] = self.data[idx][targetChannel]

        # apply window
        if applyWindow:
            fftWindow = dspUtil.createLookupTable(len(dataTmp), \
                dspUtil.LOOKUP_TABLE_HANNING)
            dataTmp *= fftWindow

        # autocorrelation
        result = numpy.correlate(dataTmp, dataTmp, mode = 'full', \
            old_behavior = False)
        r = result[result.size/2:] / readSize

        # find peak in AC
        xOfMax, valMax = findArrayMaximum(r, self.metaData.fs / Fmax, \
            self.metaData.fs / Fmin)
        valMax /= max(r)
        freq = self.metaData.fs / xOfMax
        periodSize = 0
        if freq > 0:
            periodSize = self.metaData.fs / freq
        t = (offset + (periodSize / 2.0))/self.metaData.fs
        if freq >= Fmin and freq <= Fmax and valMax >= voicingThreshold:
            self.metaData.arrF0['t'].append(t)
            self.metaData.arrF0['freq'].append(freq)
        else:
            # set F0 to zero if out of bounds
            self.metaData.arrF0['t'].append(t)
            self.metaData.arrF0['freq'].append(0)

        if periodSize > 10:
            offset += periodSize * progressPeriods
        else:
            offset += 10
    self.metaData.f0ok = True
    self.saveMetaData()

# ------------------------------------------------------------------------ #

def detectGlottalCycles(self,
        tolerance = 1, # tolerance (ms) by which cycles can be offset
        verbose = False
):
    # TODO: the cycle-detection algorithm is currently based on cross-
    # correlation with an ideal EGG waveform. this can lead to unwanted
    # gaps/overlaps between cycles, which is not ideal. This needs to
    # be fixed by introducing a better cycle-detection algorithm

    if verbose: logger('detecting glottal cycles')
    if self.metaData.f0ok == False:
        self.calculateF0()
    self.arrGlottalCycles = { 'offset':[], 'size':[] }
    arrOffset = []
    arrSize = []
    arrF0 = []
    dataSize = len(self.data)
    eggChannel = self.metaData.eggChannel
    recentOffset = 0
    for i in range(len(self.metaData.arrF0['t']) - 1):
        t1 = self.metaData.arrF0['t'][i]
        freq1 = self.metaData.arrF0['freq'][i]
        t2 = self.metaData.arrF0['t'][i + 1]
        freq2 = self.metaData.arrF0['freq'][i + 1]

        # interpolate between F0 readings, to consider every glottal cycle
        for j in range(int(self.metaData.F0progressPeriods)):
            try:
                offset = float(j) / self.metaData.F0progressPeriods
                t = interpolateLinear(t1, t2, offset)
                freq = interpolateLinear(freq1, freq2, offset)
                if freq > 0:
                    size = self.metaData.fs / freq
                    idealWaveform = getIdealEggWaveform(size)
                    offset = self.metaData.fs * t
                    left = offset - size
                    right = offset + size
                    if left < 0: left = 0
                    if right >= dataSize: right = dataSize - 1
                    dataTmp = numpy.zeros(right - left)
```

```
                    if self.metaData.channels == 1:
                        for k in range(int(right - left)):
                            dataTmp[k] = self.data[left + k]
                    else:
                        for k in range(int(right - left)):
                            dataTmp[k] = self.data[left + k][eggChannel]
                    valMax = max(dataTmp)
                    valMin = min(dataTmp)
                    dataTmp -= valMin
                    dataTmp /= (valMax - valMin)
                    xOfMax, r = corr(dataTmp, idealWaveform.data)
                    xOfMax *= -1.0
                    if xOfMax > 0:
                        xOfMax -= size * 0.1
                        if xOfMax + left > recentOffset:
                            arrOffset.append(left + xOfMax)
                            arrSize.append(size)
                            arrF0.append(freq)
                        else:
                            logger("WARNING: negative xOfMax value (" \
                                + str(xOfMax) + ") at t = " + str(t))
                    else:
                        logger("WARNING: no maximum found in correlation function")
                    recentOffset = left
                else:
                    if freq < 0:
                        logger('WARNING: F0 is negative (' + str(freq) \
                            + ') at t = ' + str(t))
            except Exception as e:
                logger(e.__str__())

        # convert offsets and sizes to integers
        numCycles = len(arrOffset)
        for i in range(numCycles):
            arrOffset[i] = int(round(arrOffset[i]))
            arrSize[i] = int(round(arrSize[i]))

        # clean cases where we have a cycle overlap, or a little gap between
        # cycles
        toleranceSamples = int(round(((tolerance / 1000.0) * self.metaData.fs)))
        for i in range(numCycles - 1):
            offset1 = arrOffset[i]
            offset2 = arrOffset[i + 1]
            size1 = arrSize[i]
            size2 = arrSize[i + 1]
            gap = offset2 - (offset1 + size1)
            if gap <> 0 and abs(gap) <= toleranceSamples:
                # gap/offset is within tolerance limits. fix it
                correction = gap / 2
                arrSize[i] += correction
                arrOffset[i + 1] -= gap - correction
                arrSize[i + 1] += gap - correction

        self.metaData.arrGlottalCycles = { 'offset':arrOffset, 'size':arrSize, \
            'F0':arrF0 }
        self.metaData.glottalCyclesOk = True
        self.saveMetaData()

    # ------------------------------------------------------------------------ #

    def getGlottalCycleRecursive(self, t, idxMin, idxMax, idxRecent, \
            tolerance, iteration
    ):
        if idxMax < idxMin:
            return -1
        idx = int((idxMax + idxMin) / 2.0)
        if idx == idxRecent:
            return -1
        offset = float(self.metaData.arrGlottalCycles['offset'][idx]) \
            / float(self.metaData.fs)
        size = self.metaData.arrGlottalCycles['size'][idx]
        duration = size / float(self.metaData.fs)
        if t + tolerance >= offset and t - tolerance <= (offset + duration):
            # found a cycle
            return idx
        if t < offset:
            # look in first half of array...
            return self.getGlottalCycleRecursive(t, idxMin, idx, idx, tolerance, iteration + 1)
        else:
```

```
        # look in second half of array...
        return self.getGlottalCycleRecursive(t, idx, idxMax, idx, tolerance, iteration + 1)

# ------------------------------------------------------------------------ #

def getGlottalCycle(
        self, t, derivative = False,
        tolerance = 0.00, normalize = True
):
    if t < 0 or t > self.metaData.duration:
        return None #glottalCycle(numpy.array([1]), self.metaData.fs)
    fs = float(self.metaData.fs)
    candidate = self.getGlottalCycleRecursive(t, 0, \
        len(self.metaData.arrGlottalCycles['offset']) - 1, -1, tolerance, 1)

    if candidate <> -1:
        try:
            size = int(self.metaData.arrGlottalCycles['size'][candidate])
            offset = float(self.metaData.arrGlottalCycles['offset'][candidate])
            dataTmp = numpy.zeros(size)
            if self.metaData.channels > 1:
                for j in range(size):
                    dataTmp[j] = self.data[j + offset][self.metaData.eggChannel]
            else:
                for j in range(size):
                    dataTmp[j] = self.data[j + offset]
            if derivative:
                dataTmp2 = numpy.copy(dataTmp)
                for j in range(size - 1):
                    dataTmp[j] = dataTmp2[j + 1] - dataTmp2[j]
                dataTmp[size - 1] = 0
                valMin = min(dataTmp)
                valMin *= -1.0
                valMax = max(dataTmp)
                for j in range(size):
                    val = dataTmp[j]
                    if val > 0:
                        val /= valMax
                    else:
                        val /= valMin
                    dataTmp[j] = val
            return glottalCycle(dataTmp, self.metaData.fs, \
                normalize = normalize)
        except Exception as e:
            print "WARNING:", e
            return None

    return None #glottalCycle(numpy.zeros(1), self.metaData.fs)
    #raise Exception('no data found')

# ------------------------------------------------------------------------ #

def createWavegram(
        self,
        width = 600, # resolution on x axis
        height = 300, # resolution on y axis
        derivative = True,
        tStart = 0,
        tEnd = -1,
        doColorize = True,
        eggDbMin = 0,
        eggDbMax = 0,
        amplitudeData = None,
        col1 = 0xFF0000,
        col2 = 0x0000FF,
        verbose = False,
        cycleDetectionTolerance = 0.01
):
    if verbose: logger('creating wavegram')
    if tEnd == -1: tEnd = self.metaData.duration
    arraySize = width * height * 3
    tmp = numpy.zeros(arraySize)
    data = tmp.reshape(height,width,3) # y, x, RGB
    if doColorize and not amplitudeData:
        amplitudeData = self.calculateRmsVector(0.1, 0.02, \
            self.metaData.eggChannel, True)
    if eggDbMin == 0 and eggDbMax == 0:
        eggDbMin = amplitudeData[1].min()
        eggDbMax = amplitudeData[1].max()
```

```
        for x in range(width):
            t = tStart + x * (tEnd - tStart) / float(width - 1)
            glottalCycle = self.getGlottalCycle(t, derivative = derivative, \
                tolerance = cycleDetectionTolerance)
            amp = 0
            found = False
            for i in range(len(amplitudeData[0])):
                if amplitudeData[0][i] >= t:
                    if i == 0: amp = amplitudeData[1][i]
                    elif i >= len(amplitudeData[0]): amp = amplitudeData[1][-1]
                    else:
                        weighting = float(t - amplitudeData[0][i - 1])
                        weighting /= float(amplitudeData[0][i] - \
                            amplitudeData[0][i - 1])
                        amp = interpolateLinear(amplitudeData[1][i - 1], \
                            amplitudeData[1][i], weighting)
                    found = True
                    break
            if not found:
                if t < amplitudeData[0][0]:
                    amp = amplitudeData[1][0]
                else:
                    amp = amplitudeData[1][-1]
            ampScaling = (amp - eggDbMin) / (eggDbMax - eggDbMin)
            #print x, amp, ampScaling
            redFg = (col1 & 0xFF0000) * ampScaling + (col2 & 0xFF0000) \
                * (1.0 - ampScaling)
            greenFg = (col1 & 0x00FF00) * ampScaling + (col2 & 0x00FF00) \
                * (1.0 - ampScaling)
            blueFg = (col1 & 0x0000FF) * ampScaling + (col2 & 0x0000FF) \
                * (1.0 - ampScaling)
            redFg /= (65536 * 256.0)
            greenFg /= (256 * 256.0)
            blueFg /= 256.0
            redBg = 1.0
            greenBg = 1.0
            blueBg = 1.0
            #print x, amp, ampScaling, redFg, greenFg, blueFg
            if glottalCycle:
                strip = wavegramStrip(glottalCycle, height)
                for y in range(height):
                    if glottalCycle.getSize() <= 1:
                        # draw white if invalid glottal cycle
                        for i in range(3):
                            data[y][x][i] = 1
                    else:
                        if doColorize:
                            val = strip.data[y]
                            r = redBg * (1.0 - val) + redFg * val
                            b = blueBg * (1.0 - val) + greenFg * val
                            g = greenBg * (1.0 - val) + blueFg * val
                            # check for clipping:
                            if r > 1: r = 1
                            if r < 0: r = 0
                            if b > 1: b = 1
                            if b < 0: b = 0
                            if g > 1: g = 1
                            if g < 0: g = 0
                            data[y][x][0] = r
                            data[y][x][1] = b
                            data[y][x][2] = g
                        else:
                            col = 1.0 - strip.data[y]
                            for i in range(3):
                                data[y][x][i] = col
            else:
                # draw white if invalid glottal cycle
                for y in range(height):
                    for i in range(3):
                        data[y][x][i] = 1
        return data

    # ------------------------------------------------------------------- #

    def saveMetaData(self):
        f = open(self.fileName + META_DATA_SUFFIX, "w")
        pickle.dump(self.metaData, f)
        f.close()
```

```
    # ------------------------------------------------------------------------ #

    def loadMetaData(self):
        f = open(self.fileName + META_DATA_SUFFIX, "r")
        self.metaData = pickle.load(f)
        f.close()

    # ------------------------------------------------------------------------ #

    def calculateRmsVector(self,
            windowSize, # seconds
            timeStep, # seconds
            channel = 0,
            convertToDb = False
        ):
        if timeStep <= 0:
            raise Exception("time step must be greater than zero")
        arrDataX = []
        arrDataY = []
        offset = 0
        while offset < self.metaData.duration:
            arrDataX.append(offset)
            tStart = offset - windowSize / 2.0
            if tStart < 0: tStart = 0
            tEnd = offset + windowSize / 2.0
            if tEnd > self.metaData.duration: tEnd = self.metaData.duration
            val = -99
            try:
                val = self.calculateRmsScalar(channel, tStart, tEnd, convertToDb)
            except Exception as e:
                print "\t\tWARNING calculateRmsVector(...):", e
            arrDataY.append(val)
            offset += timeStep
        return numpy.array(arrDataX), numpy.array(arrDataY)

    # ------------------------------------------------------------------------ #

    def calculateRmsScalar(self, channel, tStart = 0, tEnd = 0,
            convertToDb = False
    ):
        if tEnd == 0: tEnd = self.metaData.duration
        if tStart < 0:
            raise Exception("tStart must not be smaller than zero")
        if tEnd > self.metaData.duration:
            raise Exception("tEnd out of range")
        if (tEnd <= tStart):
            raise Exception("tEnd must be greater than tStart")

        #print channel, tStart, tEnd, convertToDb
        sum = 0
        offset1 = int(tStart * float(self.metaData.fs))
        offset2 = int(tEnd * float(self.metaData.fs))
        if offset1 < 0: offset1 = 0
        if offset2 >= self.metaData.numFrames:
            offset2 = self.metaData.numFrames - 1
        numSamples = offset2 - offset1
        #print offset1, offset2, channel, len(self.data[channel])
        for i in range(numSamples):
            sum += self.data[i + offset1][channel]
        mean = sum / float(numSamples)
        tmp = 0
        for i in range(numSamples):
            tmp2 = self.data[i + offset1][channel] - mean
            tmp += (tmp2 * tmp2)
        tmp /= float(numSamples + 1)
        val = math.sqrt(tmp)
        if convertToDb:
            if val == 0:
                val = -99
            else:
                val = rmsToDb(val)
        return val


######################################################################
# glottal cycle
######################################################################

class glottalCycle:
```

```
    """ this class represents sample data for one glottal cycle """

    def __init__(self, data, samplingRate, normalize = True):
        """ initialization of a glottalCycle object. unless you explicitly
            pass some data, this function is no substitute for the init(...)
            function
        """
        self.init(data, samplingRate, normalize = normalize)

    # ------------------------------------------------------------------------ #

    def getSize(self):
        return self.data.getSize()

    # ------------------------------------------------------------------------ #

    def init(self, data, samplingRate, normalize = True):
        """ data is stored in a normalized form (y = [0..1]) """
        # make sure data is of right type
        objType = type(data).__name__.strip()
        if objType == "ndarray":
            self.data = wavegramArray(data) # cast to wavegram.wavegramArray
            self.size = len(data)
        elif objType == "list":
            self.data = wavegramArray(data) # cast to wavegram.wavegramArray
            self.size = len(data)
        else:
            raise Exception('data argument is no instance of numpy.array')

        # set member variables
        self.samplingRate = samplingRate
        self.f0 = self.samplingRate / float(self.size)

        #normalize
        valMax = self.data.data.max()
        valMin = self.data.data.min()
        for i in range(self.size):
            if normalize:
                self.data.data[i] = (self.data.data[i] - valMin) \
                    / (valMax - valMin)



########################################################################
# a wavegram "strip"
########################################################################

class wavegramStrip:
    """ this class represents one "strip" in the wavegram. a strip can be
        created by passing an object of type glottalCycle to the init function
    """

    def __init__(
            self,
            data,
            size,
            normalize = True # if False, a non-normalized version is created
    ):
        self.init(data, size, normalize)

    # ------------------------------------------------------------------------ #

    def init(
            self,
            data,
            size,
            normalize = True # if False, a non-normalized version is created
    ):

        if isinstance(data, glottalCycle):
            if normalize:
                dataTmp = data.data
                dataTmp.stretch(size)
                self.data = dataTmp.data
            else:
                self.data = data.data
        else:
            raise Exception('expected object of class glottalCycle')
        self.normalized = normalize
```

Supplement E: Curriculum Vitae

Name:	Christian Thomas HERBST, M. A.
Date and place of birth:	Dec 5, 1970, Salzburg/Austria
Nationality/Citizenship:	Austrian
Children:	four (born 1993, 2001, 2003 and 2009)
Email:	herbst@ccrma.stanford.edu
Web:	www.christian-herbst.org

Education

2008 - ongoing:	PhD-Student (supervisor Dr. Jan Švec), Palacký University Olomouc, Faculty of Sciences, Department of Biophysics
1997 – 2005:	Masters degree in music pedagogy (voice) / applied vocology from the University Mozarteum Salzburg; voice coaching by Dr. Albert Hartinger
1990 – 1992:	Studies in opera singing at the University Mozarteum Salzburg with Eva Trefas-Illes
May 1990:	Abiture from the Commercial College in Salzburg

Educational/Research Stays

Sep. 2009 – present:	Bioacoustics Lab, Dept. of Cognitive Biology (W. T. Fitch), University Vienna/Austria
Feb. 2006:	Guest Researcher at the Groningen Voice Research Lab, University of Groningen, the Netherlands
March – Sep. 2004:	Erasmus Exchange Student at the Department of Speech, Music and Hearing of the KTH Stockholm / Sweden,
Summer 2001:	Summer Vocology Institute, organized by the National Center for Voice and Speech and the University of Iowa (Department of Speech Pathology and Audiology). Courses in Principles of Voice Production, Instrumentation for Voice Analysis and Voice Habilitation
Sep. 1999 – May 2000:	Visiting researcher at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University

Professional Experience (selected)

Sep. 2009 – present:	Bioacoustics Lab, Dept. of Cognitive Biology (W. T. Fitch), University Vienna/Austria
March 2009 – present:	Consultation of voice patients, ENT clinic Dr. Josef Schlömicher-Thier, Neumarkt am Wallersee, Austria

Oct. 2009 – present:	University Mozarteum Salzburg: Part-time lecturer “Anatomie, Physiologie, Körperschulung 01 und 02” and “Bewegungsphysiologie und Musikermedizin”
Feb. 2006 – Oct. 2009:	Voice teacher, assistant choir director and scientific consultant of the Tölzer Knabenchor, Munich/Germany
Nov. 2001 – June 2007:	Independent software engineer and consultant
Sep. 2004 – Jan. 2006:	Part-time teacher (Voice Coaching for Choirs) at the Musikum Salzburg
Sep. 2002 – Dec. 2005:	Voice pedagogue of the Salzburg Cathedral Youth Choir
March 2003 – June 2003:	Part-time lecturer at the Institute for Stage Design, University Mozarteum Salzburg

Reviewing

2011	J Voice (Resident Editor), J Acoust Soc Am, Current Bioinformatics
2010	J Voice
2008	Logopedics Phoniatrics Vocology

Membership in professional organizations

2010 - present	British Voice Association
2009 - present	COMET (Collegium Medicorum Theatri)

Grants, Scholarships and Awards

G5. Dean's Prize. Palacký University Olomouc, Faculty of Science, December 2011. Awarded for the publication Herbst CT, Qiu Q., Schutte HK, Švec JG: Membranous and cartilaginous vocal fold adduction in singing. J Acoust Soc Am 129(4): 2253-2262 (2011).

G4. Dean's Prize. Palacký University Olomouc, Faculty of Science, May 2011 (awarded for the best PhD paper in Physics)

G3. Research Grant. Salzburger Landesregierung (Salzburg County Government), June 2006

G2. SEMPRE Conference Award. Society for Education, Music and Psychology Research, May 2006

G1. Promotion Grant ("Förderstipendium"). University Mozarteum, December 2004

Supplement F: Publications by the author

Peer-reviewed Articles

A7. Herbst, C. T., Howard, D. M., & Švec, J. G. (submitted). The sound source in singing – basic principles and muscular adjustments for fine-tuning vocal timbre. In G. Welch, D. M. Howard & J. Nix (Eds.), The Oxford Handbook of Singing. Oxford, UK: Oxford University Press.

A6. Christian T. Herbst, Qingjun Qiu, Harm K. Schutte, Jan G. Švec (2011). Membranous and cartilaginous vocal fold adduction in singing. J Acoust Soc Am, 129 (4), 2253-2262

A5. Christian T. Herbst, W. T. S. Fitch, Jan G. Švec (2010). Electroglottographic wavegrams: a technique for visualizing vocal fold dynamics noninvasively. J Acoust Soc Am, 128 (5), 3070-3078

A4. Christian T. Herbst, David Howard, Josef Schlömicher-Thier (2010). Using electroglottographic real-time feedback to control posterior glottal adduction during phonation. J Voice, 24 (1), 72 - 85

A3. Christian T. Herbst, Jan G. Švec, Sten Ternström (2009). Investigation of four distinct glottal configurations in classical singing - a pilot study. J Acoust Soc Am – Express Letters, 125 (3), EL104-EL109

A2. Christian T. Herbst (2007). Der Knabensolist in der Oper - Ein akustisches Portrait. L.O.G.O.S. Interdisziplinär, 15 (3), 166-174

A1. Christian T. Herbst, Sten Ternström (2006). A comparison of different methods to measure the EGG contact quotient. Logopedics Phoniatrics Vocology, 31 (3), 126-138

Posters (presented at conferences)

P4. Christian T. Herbst, W. T. S. Fitch, Jan G. Švec (2010). Wavegrams: A new technique for visualizing vocal fold dynamics noninvasively using electroglottographic signals. COST Action 2103 Summer School - Modeling and Assessment of the Human Voice, Erlangen, Germany. September 2010. (http://www.christian-herbst.org/poster_wavegrams.pdf)

P3. Josef Schlömicher-Thier, Donald G. Miller, Hubert Noe, Christian T. Herbst (2009). Yodeling - Acoustic and Physiologic Properties. The Voice Foundation's 38th Annual Symposium, June 2009. (http://www.christian-herbst.org/poster_yodelling.pdf)

P2. Christian T. Herbst, Elke Duus, Harald Jers (2009). Voice category assessment of amateur choir singers. 4th International Conference on the Physiology and Acoustics of Singing, January 2009. (http://www.christian-herbst.org/poster_pas4.pdf)

P1. Christian T. Herbst, Jan G. Švec, Qingjun Qiu, Harm Schutte (2007). Overall and posterior glottal adduction in singing. 7th Pan European Voice Conference (PEVOC), Groningen, The Netherlands. August 2007. (http://www.christian-herbst.org/poster_pevoc_2007.pdf)

Conferences and Lectures (50 total, 14 invited lectures)

C50. Christian T. Herbst (2011). Angewandte Stimmphysiologie und Akustik. CAS Singstimme -- Fehlfunktionen erkennen, abbauen, vermeiden (invited lecture), Hochschule der Künste Bern, Bern, Switzerland. November 26, 2011.

C49. Jakob Unger, Tobias Meyer, Christian T. Herbst, Michael Döllinger, Paul Lohscheller (2011). PVG-Wavegramm: Dreidimensionale Visualisierung von Stimmlippendynamik. 28. Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie e. V., Zurich, Switzerland. September 10, 2011, presented by Jakob Unger.

C48. Christian T. Herbst, W. T. S. Fitch, Jan G. Švec (2011). Observing the female middle register using EGG wavegrams. 9th Pan-European Voice Conference (PEVOC) (invited lecture), September 1, 2011.

C47. Christian T. Herbst, Jan G. Švec (2011). Voice acoustics, microphones, recording and computers. One-Day Crash Course on 'Voice' (invited lecture), European Academy of Voice, August 30, 2011.

C46. Jakob Unger, Tobias Meyer, Christian T. Herbst, Michael Döllinger, Paul Lohscheller (2011). PVG-Wavegrams: Three-dimensional visualization of vocal fold dynamics. 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), August 25, 2011, presented by Jakob Unger.

C45. Christian T. Herbst, W. T. S. Fitch, Jan G. Švec (2011). Wavegrams: A new technique for visualizing vocal fold dynamics noninvasively using electroglottographic signals. 40th Annual Symposium: Care of the Professional Voice, The Voice Foundation, June 2, 2011.

C44. Christian T. Herbst (2011). Vocal folds and the voice timbre. 3rd Czech-Slovak Symposium on ART VOICE (invited lecture), Hlasové a sluchové centrum Praha, s.r.o., Prague, Czech Republic. May 21, 2011.

C43. Christian T. Herbst (2011). Wie funktioniert Stimme. Regensburger Stimmtag: ein regionaler Beitrag zum World Voice Day (invited lecture), Regensburger ärztenetz e.V., Regensburg, Germany. April 16, 2011.

C42. Christian T. Herbst (2011). Stimmbildungsunterricht mit Knaben und Mädchen. Kinderchorleitungssymposium (invited lecture), Universität der Künste, Berlin, Germany. February 11, 2011.

C41. Christian T. Herbst (2011). Beschaffenheit und Ausbaumöglichkeit der Kinderstimme. Kinderchorleitungssymposium (invited lecture), Universität der Künste, Berlin, Germany. February 11, 2011.

C40. Christian T. Herbst (2010). Understanding vocal timbre in singing - a tutorial. Europa Cantat General Assembly 2010 (invited lecture), Europa Cantat, Namur, Belgium. November 28, 2010.

C39. Christian T. Herbst, Josef Schlömicher-Thier, Matthias Weikert (2010). Berufsstimmbetreuung in der HNO-Praxis - eine stimmdiagnostische, -therapeutische und gesangsdidaktische Synopsis. Stuttgarter Stimmtage, Staatliche Hochschule für Musik und Darstellende Kust, Stuttgart, Germany. October 2, 2010.

C38. Josef Schlömicher-Thier, Hans E. Eckel, Christian T. Herbst (2010). Interventionelle Laryngologie in der HNO-Praxis. 54th annual Meeting of the Austrian Society of Oto-Rhino-Laryngology, Head and Neck Surgery, Salzburg, Austria. September 17, 2010. presented by Josef Schlömicher-Thier.

C37. Christian T. Herbst, W. T. S. Fitch, Jan G. Švec (2010). Wavegrams: A new technique for visualizing vocal fold dynamics noninvasively using electroglottographic signals. 9th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research (AQL), September 2010.

C36. Christian T. Herbst (2010). Das Timbre im klassischen Gesang: akustisches und physiologisches Tutorial. 9th Voice Symposium Salzburg, Austrian Voice Institute, Salzburg, Austria. August 28, 2010.

C35. Christian T. Herbst, W. T. S. Fitch, Jan G. Švec (2010). Visualizing electroglottographic signals with wavegrams. 5th International Conference on the Physiology and Acoustics of Singing (PAS5), Kungliga Tekniska Högskolan, Stockholm, Sweden. August 11, 2010.

C34. Christian T. Herbst, Jan G. Švec, Qingjun Qiu, Harm Schutte (2010). Membranous and cartilaginous glottal adduction in singing - experimental findings and pedagogic considerations. Choice for Voice Conference, British Voice Association, London, U. K.. July 2010.

C33. Jan G. Švec, Jaromír Horáček, Tomáš Vampola, Christian T. Herbst, Donald G. Miller, Radan Havlík, Petr Krupa, Mojmír Lejska (2010). Acoustic and articulatory adjustments in operatic singing: Spectral analysis and magnetic resonance imaging. COST 2103: 4th Advanced Voice Function Workshop AVFA'10, York, U.K.. May 20, 2010, presented by Jan G. Švec.

C32. Christian T. Herbst, Josef Schlömicher-Thier (2010). Die Sängerbetreuung in stimmpädagogischer und sängermedizinischer Kooperation. Berner Symposium Medizin, Logopädie, Gesangspädagogik (invited lecture), Hochschule der Künste Bern, Bern, Switzerland. April 17, 2010.

C31. Christian T. Herbst, Josef Schlömicher-Thier (2010). Pedagogical and Medical Cooperation in Voice Patient Care. Symposium Ars Choralis (invited lecture), Croatian Choral Directors Association, Zagreb, Croatia. April 10, 2010, presented by Christian T. Herbst.

C30. Christian T. Herbst (2010). Voice Timbre in Singing. Symposium Ars Choralis (invited lecture), Croatian Choral Directors Association, Zagreb, Croatia. April 10, 2010.

C29. Christian T. Herbst (2010). Vocal Fold Adduction and Registers in Classical Singing. Symposium Ars Choralis (invited lecture), Croatian Choral Directors Association, Zagreb, Croatia. April 8, 2010.

C28. Ramona Steiner, Christian T. Herbst, David Howard (2009). Electroglottographic (EGG) real-time biofeedback to enhance glottal adduction in patients with unilateral vocal fold pareses. 8th Pan European Voice Conference (PEVOC), Dresden, Germany. August 28, 2009, presented by Ramona Steiner.

C27. Christian T. Herbst, Jan G. Švec, Qingjun Qiu, Harm Schutte (2009). Membranous and cartilaginous glottal adduction in singing. 8th Pan European Voice Conference (PEVOC), Dresden, Germany. August 27, 2009.

C26. Jan G. Švec, Jaromír Horáček, Tomáš Vampola, Christian T. Herbst, Donald G. Miller, Radan Havlík, Petr Krupa, Mojmír Lejska (2009). Acoustic and Articulatory Adjustments for Singers' Formant Production: Spectral Analysis, MRI and Finite Element Modeling. The Voice Foundation's 38th Annual Symposium, The Voice Foundation, Philadelphia. June 2009, presented by Jan G. Švec.

C25. Jan G. Švec, Christian T. Herbst, Sten Ternström (2009). Membranous versus cartilaginous glottal adduction in four singing voice qualities: Pilot laryngostroboscopic and

videokymographic observations. Proceedings of AVFA '09, 3rd Advanced Voice Function Assessment International Workshop, May 18, 2009, presented by Jan G. Švec.

C24. Christian T. Herbst, Josef Schlömicher-Thier (2009). Stimme - Ausdrucksmittel und Werkzeug im Kunstbetrieb. Symposium: Internationales Theaterinstitut der UNESCO - Centrum österreich, Vienna, Austria. March 28, 2009.

C23. Jan G. Švec, Christian T. Herbst, Radan Havlík, Jaromír Horáček, Petr Krupa, Mojmír Lejska, Donald G. Miller (2008). Singer's formant: Preliminary results of MRI and acoustic evaluations of singers. Proceedings Interaction and Feedbacks 2008, Prague: Institute of Thermomechanics AS CR, November 2008, presented by Jan G. Švec.

C22. Christian T. Herbst, Elke Duus (2008). Stimmliche Leistungsbeurteilung von SängerInnen im Amateurchor. 8th Voice Symposium Salzburg, Austrian Voice Institute, Salzburg, Austria. July 27, 2008.

C21. Christian T. Herbst (2008). Pressen, behauchtes Singen und Registerdivergenzen - Einfluss der glottischen Konfiguration auf das Timbre im klassischen Gesang. 8th Voice Symposium Salzburg, Austrian Voice Institute, Salzburg, Austria. July 26, 2008.

C20. Christian T. Herbst (2008). Einfluss der Stimmlippentätigkeit auf das Gesangstimbre. Guest Lecture, Hochschule für Musik, Köln, June 11, 2008.

C19. Christian T. Herbst, Josef Schlömicher-Thier (2007). Visualization and Analysis of Electroglottographic Waveforms. XVI Annual PVSF/UCLA Voice Conference, Los Angeles, CA. October 25, 2007, presented by Josef Schlömicher-Thier.

C18. Christian T. Herbst (2007). Glottal Contact in Singing. 5th international logopedics and phoniatrics course 'THE ARTISTIC VOICE' (invited lecture), La Voce Artistica, Ravenna, October 18, 2007.

C17. Christian T. Herbst (2007). Kehlkopfkonfigurationen beim Singen. Symposium: Stimmbildung in Knaben-, Mädchen- und gemischten Kinderchören (invited lecture), Universität der Künste Berlin, Germany, October 4, 2007.

C16. Christian T. Herbst, Jan G. Švec (2007). Is the degree of posterior glottal adduction relevant for "voix mixte" phonation? 7th Pan European Voice Conference (PEVOC), Groningen, The Netherlands. September 1, 2007.

C15. Christian T. Herbst, David Howard, Josef Schlömicher-Thier (2007). Using electroglottographic real-time feedback to control posterior glottal adduction during phonation. 7th Pan European Voice Conference (PEVOC), Groningen, The Netherlands. September 1, 2007.

C14. Christian T. Herbst (2007). Der Knabensolist in der Oper - Akustisches Portrait eines musikalischen Hochleistungssportlers. 75. Kongress der Deutschen Gesellschaft für Sprach- und Stimmheilkunde (DGSS), April 21, 2007.

C13. Gerhard Schmidt-Gaden, Christian T. Herbst, Ernst L. Schmid (2007). Talentschmiede Knabenchor? 19. Jahreskongress des Bundesverbandes Deutscher Gesangspädagogen (BDG), April 20, 2007.

C12. Christian T. Herbst, Josef Schlömicher-Thier (2007). Stimmbildung und Stimmstörungsprävention. Internationale Tagung "Die Stimme Heute", Zentralkrankenhaus Bozen, Abteilung HNO, January 26, 2007, presented by Christian T. Herbst.

C11. Christian T. Herbst (2006). Stimmbandschluss im klassischen Gesang. Grazer Stimmtage, Hals-, Nasen-, Ohren-Universitätsklinik Graz, Klinische Abteilung für Phoniatrie, November 25, 2006.

C10. Christian T. Herbst (2006). Physiologische Vorgänge beim Registerausgleich der Knabenstimme. 1. Internationales Symposium für Kinderstimmbildung, Freunde der Wiener Sängerknaben, November 4, 2006.

C9. Christian T. Herbst (2006). Acoustic Priciples of Voice Production - a Tutorial. 7th Voice Symposium Salzburg, Austrian Voice Institute, August 4, 2006.

C8. Josef Schlömicher-Thier, Phillip Janssen, Christian T. Herbst (2006). Phonetogram: Architecture of Speaking and Singing Voice. 3rd World Voice Conference, Istanbul, Turkey. June 22, 2006.

C7. Christian T. Herbst, Josef Schlömicher-Thier, Matthias Weikert (2006). Voice Disorders in Childhood & Management of Mutational Problems in Choirboys. 3rd World Voice Conference, Istanbul, Turkey. June 21, 2006.

C6. Christian T. Herbst, Jan G. Švec (2006). Investigation of four distinct glottal configurations in aclassically trained male singer. 3rd physiology and acoustics of singing conference (PAS3-06), University of York, United Kingdom. May 11, 2006.

C5. Christian T. Herbst (2006). Untersuchung der Sängerstimme mittels Elektroglottographie (Workshop). 6. Wiener gesangswissenschaftliche Tagung, Institut Antonio Salieri, Universität für Musik und darstellende Kunst, Wien, Vienna, Austria. January 14, 2006.

C4. Christian T. Herbst (2005). The Singer's Voice: Glottal Configurations and Voice Source Properties. Seminar, School of Arts, Culture & Environment, University Edinburgh, Edinburgh, UK. October 27, 2005.

C3. Christian T. Herbst, Sten Ternström (2005). A comparison of different methods for measuring the electroglottographic contact quotient. 6th Pan European Voice Conference (PEVOC), London, UK. September 2, 2005.

C2. Christian T. Herbst (2005). Die Singstimme als physikalisch-akustisches System. Seminar, Musikum Salzburg, Salzburg, Austria. January 22, 2005.

C1. Christian T. Herbst (2004). The EGG Contact Quotient as a Means of Assessing Vocal Registration Quality in Classical Singing. Seminar, Dept. of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Sweden. August 24, 2004.

Radio Interviews

I6. Paul Lohberger (2011). Radiokolleg - Die Stimme als Instrument. Österreichischer Runkfunk Ö1, September 5 - 8, 2011

I5. Paul Lohberger (2011). Radiodoktor - Das Ö1 Gesundheitsmagazin: Stimmbildung - Therapie und Körpererfahrung. Österreichischer Runkfunk Ö1, March 23, 2011

I4. Katrin Müller-Höcker (2010). Vibrierende Muskeln, klingender Atem - Wie funktioniert das Wunderwerk Stimme. Bayern 2, November 28, 2010

I3. Paul Lohberger (2010). Radiokolleg - Queen. Österreichischer Runkfunk Ö1, May 4, 2010

I2. Paul Lohberger (2009). Radiokolleg - Die Stimme. Österreichischer Runkfunk Ö1, April 29, 2009

I1. Bayerischer Rundfunk (2008). Engelsgleich - Über die Physik der Knabenstimme. Bayern 4 Klassik, April 30, 2008