A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences

Leonardo Fuks[1], Britta Hammarberg[2] and Johan Sundberg

Abstract

This investigation describes various characteristics of a particular phonation mode, vocal-ventricular mode (VVM), as produced by a healthy, musically-trained subject. This phonation mode was judged as perceptually identical to that used in the Tibetan chant tradition. VVM covered a range close to an octave, starting at about 50 Hz. High-speed glottography revealed that the ventricular folds oscillated at half the frequency of the vocal folds thus yielding a frequency of $f_0/2$. Phonation at $f_0/3$ was also possible. Presumably, aerodynamic forces produced by the glottal flow pulses sustained the vibrations of the ventricular folds. Complementary aspects of this type of phonation were compared to phonation in modal and pulse registers by acoustical analysis of the audio signal, by inverse filtering of the flow signal and by electroglottography (EGG). In addition, oesophageal pressures were measured. These analyses revealed that every second flow pulse was attenuated because of the ventricular fold vibrations and that the laryngeal contact area alternated between two minimum values. The spectra of VVM sounds contained clear harmonic partials up to about 4 kHz. Oesophageal pressure tended to change when phonation switched between VVM and modal phonation. Examples of periodic pulse register, another case of voice period multiplication, produced patterns of EGG waveform differing from those of VVM. The possibilities of using VVM in contemporary music, whether in a purely vocal form or during to wind instrument playing, are discussed.

Introduction

In certain types of voice pathology the ventricular folds (Figure 1) vibrate (Freud, 1962; Boone, 1983; Titze, 1994). According to Titze (1994), such vibrations also occur in vocal effects used in some pop singing; for instance, it has been assumed that Louis Armstrong used the ventricular folds in singing. It has been reported that the false vocal folds may adduct more than the true ones, vibrating independently and in other cases are adducted together, establishing a complicated vibration pattern. In cases of pathological phonation, the ventricular oscillations are often non-periodic. However, periodic oscillation has been documented in patients with ventricular dysphonia, e.g. by Nasri and co-workers (1996). In trained classical western singers, on the other hand, the ventricular folds are described to flatten against the lateral wall of the larynx (Pressman & Kelemen, 1955). Studying Tuvanian performers of overtone singing by means of x-ray photography, Dmitriev et al. (1983) observed a close approximation of the ventricular folds. Lindestad and co-workers (1998) have recently reported simultaneous oscillations of the ventricular and vocal folds, observed by high-speed glottography, on a trained subject singing in the Mongolian style.

Ventricular oscillations generally occur at lower frequencies than those normally observed in the vocal folds. However, other types of phonation may also present low frequencies without necessarily involving ventricular fold activity. Pulse register refers to a mode of phonation with a pulse-like vibratory pattern (Hollien, 1974). It is generated in the glottis and sometimes also referred to as vocal fry, glottal fry, creak, creaky voice and Strohbass. Pulse register does not even need to be periodic. Regarding Strohbass, used in some ethnical types of music (e.g., "Don-Cossack" singing),

[1] UFRJ- Rio de Janeiro Federal University-School of Music, Rua do Passeio 98, 20021-290, Brazil

[2] Also at Karolinska Institute, Dept. of Logopedics and Phoniatrics, Huddinge Univ. Hosp., SE-14186, Huddinge, Sweden

regularity is probably mandatory. A periodic pulse register may sound lower than 20 Hz in some cases.

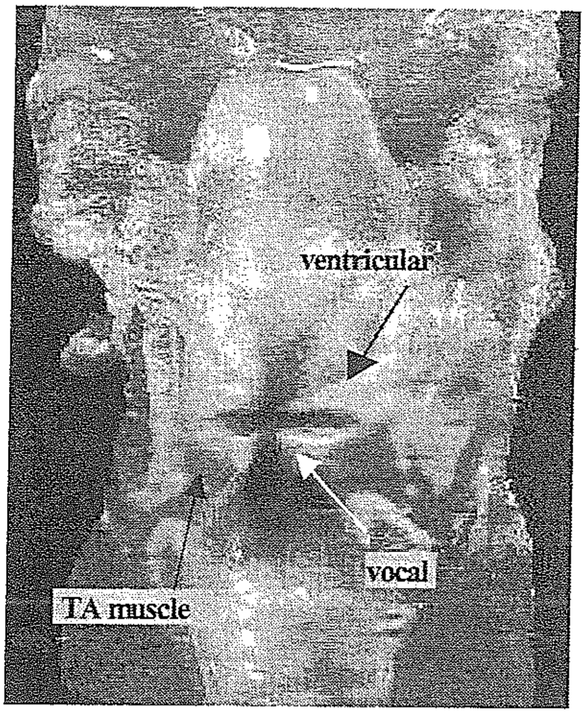

Figure 1. A coronal section through a dissected larynx, showing the vocal and ventricular folds and the thyroarythenoid muscle (TA), which includes the vocalis muscle. The configuration in this picture should be close to that during vocal-ventricular phonation. Photo adapted from Rohen & Yokochi (1993).

The abundant citation of ventricular fold vibration in some types of phonation still seems to be lacking a comprehensive theory, and experimental data are apparently scarce. The aim of the present investigation was to apply a combination of presently available analysis techniques to periodic ventricular fold phonation in order to elucidate how such vibrations are generated. Also, this kind of phonation was compared to pulse and modal register phonation.

Materials and methods

The subject, co-author LF, is a professional wind-instrument musician and acoustician at the age of 36 years who also has experience in several extended vocal techniques (Barnett, 1977). The subject phonated with a particular voice effect that appeared related to ventricular fold vibration. Preliminary laryngeal video-stroboscopy had been performed under similar conditions, indicating simultaneous vibration of the vocal folds and the ventricular folds. However, the degree of resolution and the frame rate achieved did not allow for information on phase relationships and more detailed representation of the whole cycle. The subject sustained a low-pitched tone (Bb_1, approx. 58 Hz) on the vowel /ɞ/. The Bb_1 was chosen due to the fact that it was rather comfortable to produce and offered a SPL range of approximately 64 dB (with closed mouth) to 88 dBA at 0.3 m distance. Any vowel could be produced within the range G_1 to $F\#_2$, approximately. The sound was generally initiated one octave higher (Bb_2) and then suddenly switched to the lower frequency.

In addition to the steady tones, the subject used an overtone-singing technique on the same effect, producing the melody of "Oh Susanah" on a fixed fundamental frequency of 53 Hz, approximately. Also, the subject produced examples of periodic pulse register, alternating with modal register at the pitches of $C\#_3$ (138 Hz) and $G\#_3$ (207 Hz). The pulse register also produced the perceived pitches of $f_0/2$ and $f_0/3$, for $C\#_3$ and $G\#_3$, respectively. These examples were recorded in the same way as described below, with the exception of high-speed photography and inverse-filtering.

Audio, videolaryngo-stroboscopic, high-speed video imaging, electroglottographic and sub-glottal pressure recordings were performed. A Speedcam+ high-speed camera (Weinberger AG) with a resolution of 256x64 dpi was employed for obtaining the high-speed glottograms (Eysholdt et al., 1994). The recording was carried out at the Huddinge Hospital, department of logopedics and phoniatrics. The frame rate was 1904 per second, and each take lasted for approximately one second. The sequence and the audio signals were digitally stored on a PC and also on a video tape. For laryngoscopy, a rigid telescope (70° Hopkins 8706 CJ, Karl Storz) with a light source (Olympus Xenon, type CLV-U20) was employed. Several takes were recorded and stored.

On a different occasion, the transglottal airflow during this type of phonation was analysed by inverse-filtering (Rothenberg, 1973). The flow signal was captured by a Rothenberg mask, while the subject sustained the same pitch and vowel as during the video-recorded session. Flow calibration was performed at 400 and 1200 l/h, obtained from an air bottle attached to a flow meter. The audio signals were captured by the microphone of a sound-level meter (Ono Sokki LA-210) that was placed at 1 m distance, the output of which also contained the SPL readings. The recordings were performed in a sound-treated room.

Electroglottography (EGG) was simultaneously accomplished by attaching the electrodes (Glottal Enterprises type SC-1B) over the thyroid cartilage area (Fabre, 1957; Orlikoff, 1998). All the above-mentioned signals were directly recorded into a PC computer, through a DT- 2821 Data Translation board.

Subglottal pressures during alternation between modal voice and the particular phonatory effect considered were indirectly measured by a thin pressure transducer (Gaeltec CTO-2 strain gauge catheter, 2 mm diameter; Gaeltec S7b amplifier) introduced through the nose and placed at approximately 4-6 cm below the upper oesophageal sphincter. Those pressure signals were calibrated against a standard manometer connected to the mouth, while the subject produced three levels of static pulmonary pressures: 0, 20 and 40 cm H_2O (Fry et al., 1952).

All data were recorded as files under the SMP format, using SoundSwell program with a sampling frequency of 16 kHz per channel. For data processing comprising amplification, post-calibration, filtering, spectral analysis and display of data, SoundSwell and Extract (Nyvalla DSP, Stockholm) software packages were used.

Subjective expert evaluation

The subject demonstrated these voice effects to a Tibetan Buddhist, who had spent his life up to mature age in a Tibetan community, and asked him to judge how similar these effects were to the traditional voice produced by certain Buddhist monks. The expert also judged the pulse register voice, produced at the same pitch as before. Examples of these two types of phonation were demonstrated in quasi-random order. He found the first one practically identical to that used by Tibetan monks, while he considered the pulse register voice as "slightly similar but not as loud and deep as the typical Tibetan sound".

Analysis and results

High-speed imaging

In all recordings, the high-speed video film showed a similar and regular pattern, revealing that the ventricular folds closed symmetrically at every second cycle of the vocal folds. This oscillation mode will henceforth be referred to as the *vocal-ventricular mode* (VVM). A short sequence, comprising about 3 cycles was converted from the videotape into a digital MPEG-compressed format, by means of a Silicon Graphics O2 workstation and using Media Recorder program. Then, 18 equally-spaced frames of a complete cycle, corresponding to approx. 17.2 milliseconds were captured. This sequence of frames is shown in Figure 2. The indicated angles refer to the 40 degrees angular steps relative to f_0, 116 Hz.

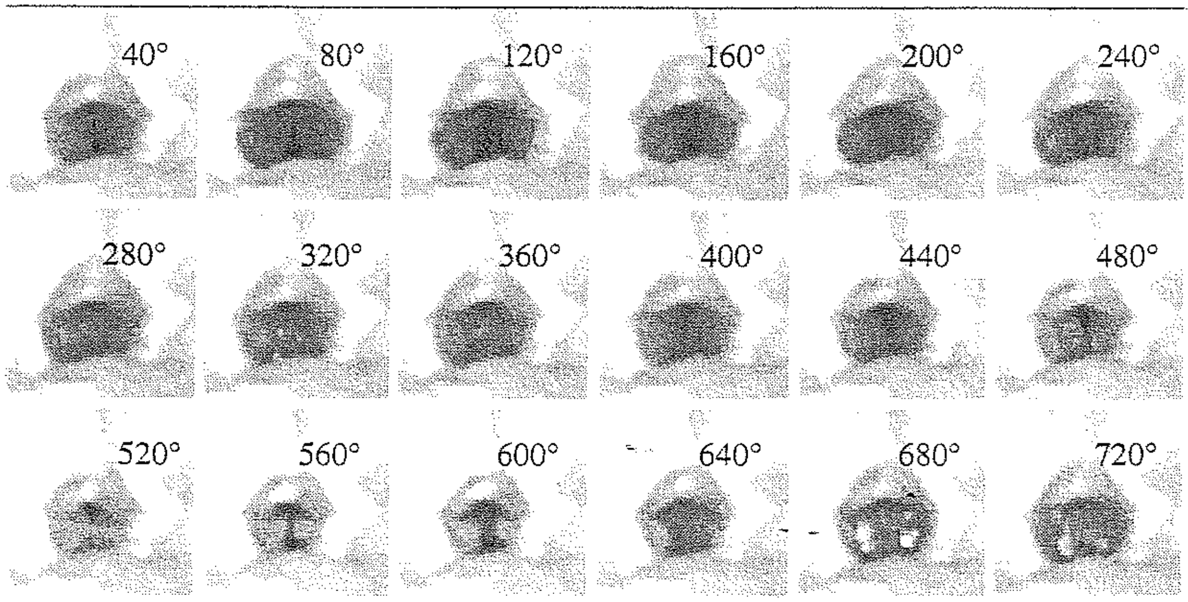

Figure 2. A complete cycle of the vocal-ventricular mode (VVM), obtained from high-speed glottography at 1904 frames/sec. Fundamental frequency of the vocal folds is 116 Hz and the ventricular folds oscillate at 58 Hz. The first cycle is purely vocal (0°-360°), while the second cycle is affected by the ventricular closure, during the 480°-560° interval, when the vocal folds were open. Ventricular folds appear as a pair of round and light-tone protuberances. Duration of the whole sequence is 17.2 milliseconds, approximately.

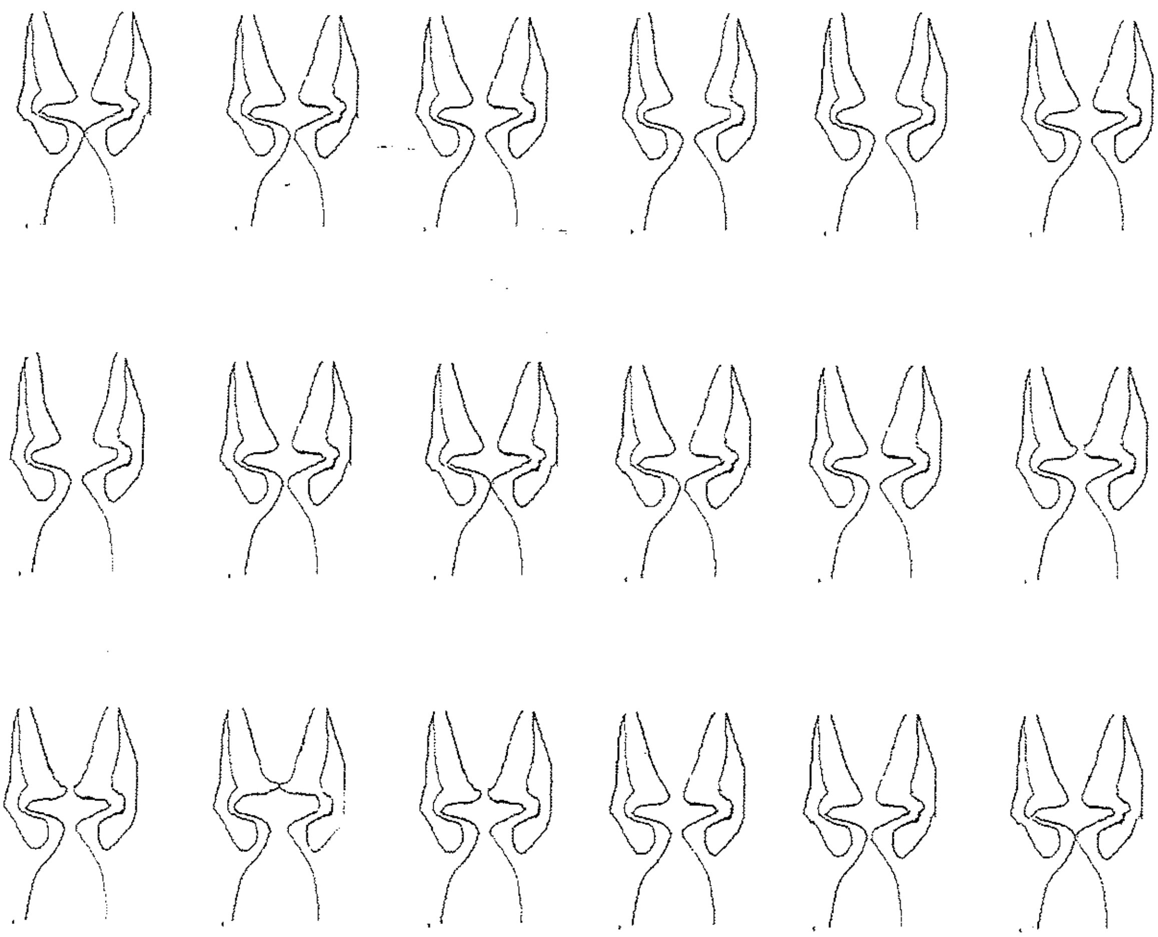

Figure 3. Sequential diagram of a whole $f_0/2$ vocal-ventricular cycle in coronal section, starting from closed glottis, see Fig. 2. Each frame corresponds to a 40° angular step relative to f_0 (116 Hz).

The images provided by the high-speed video recording display a superior view of the glottis. An attempt to construct a sequential diagram of the larynx in coronal view was done on the basis of the sequence shown in Figure 2. Figure 3 shows the result in terms of a schematic interpretation of the shapes of the folds and the distances between the structures. To gain more detailed information other imaging techniques need to be recruited.

Comparing modal register to vocal-ventricular mode (VVM)

Figures 4 and 5 show simultaneous audio waveform, flow glottogram (FG), electroglottogram (EGG) and power spectrum of modal voice and its corresponding VVM, respectively. As compared to Figure 4, presenting a sequence of four cycles of the audio signal, the second and fourth cycles in Figure 5 are damped to approximately one third of the amplitude, i.e. -10 dB. The corresponding pulses of the FG signal were even more damped, -13 dB approx. The duration between the pulse peak and the intersection with the baseline for the FG curves, i.e. the duration of the closing phase, was estimated from a series of eight consecutive cycles. Modal voice showed a longer closing duration (1.63 ms) than the VVM (1.50 ms). Furthermore, the amplitude of the flow pulse for the modal voice was considerably higher (Figures 4 and 5). This means that in VVM the descending part of the FG curves was significantly steeper. As a consequence, its spectrum can be expected to contain more energy for the higher partials. For a more profound analysis of the FG signals, a time derivative of the glottal flow curves should be obtained and the parameters which describe the waveshape calculated. To quantitatively predict the impact of these differences on

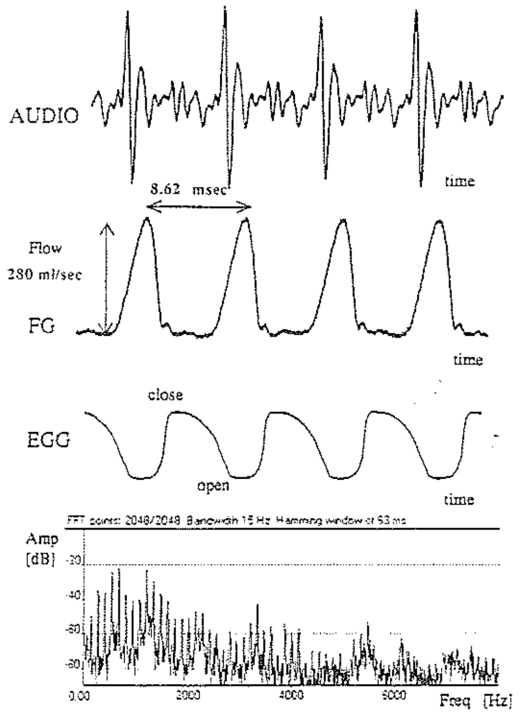

Figure 4. Audio, Inverse-filtered (FG), EGG and power spectrum representations of an /Œ/ vowel, ***modal register,*** *at 116 Hz (Bb₂). The EGG waveform shows a constant peak-to-peak pattern.*

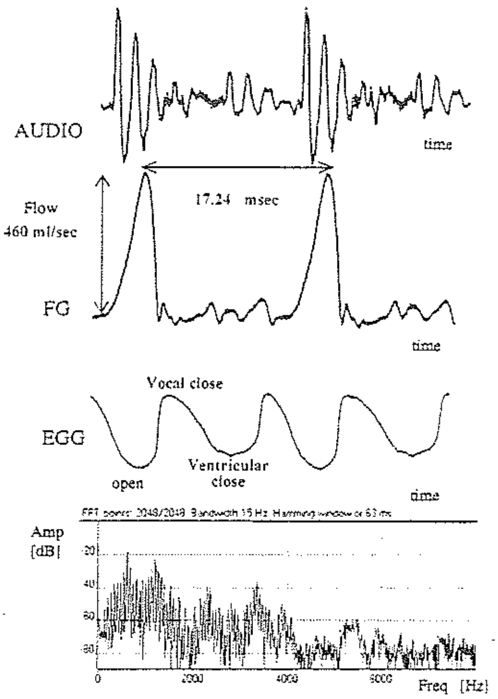

Figure 5. Audio, Inverse-filtered signal (FG), EGG and power spectrum representation of an /Œ/ vowel, VVM, vocal folds at 116 Hz (Bb_2); ventricular folds at 58 Hz (Bb_1). EGG signals peak at consistent values at every semi-cycle, but alternate in lower values between open glottis and closed ventricular folds.

formant levels the LF-model of glottal flow (Fant et al., 1985; Fant, 1997) can be applied.

The EGG signal (Figure 5) illustrates the degree of glottal contact area. The flat baseline appearing in Figure 4 (high impedance-open glottis), cannot be seen, but rather an alternation between a higher and a lower position of the valleys reflecting an alternation between two degrees of contact. The somewhat higher conductance of the larynx occurs precisely when the ventricular folds meet at every second cycle between two closures of the vocal folds. The peaks, corresponding to vocal fold closure, are similar in amplitude.

The spectra of the audio signals in Figures 4 and 5 have similar envelopes, since the same vowel was produced in both cases. Both show a harmonic spectrum, although the density of the partials for the VVM tone is twice that of the modal. However, the VVM phonation exhibits more energy between 2 kHz and 6 kHz than the modal. Figure 6 displays the same spectra in the range 0-1 kHz. In the VVM spectrum, the 58 Hz "undertone" is considerably weaker than the 116 Hz partial; generally a difference of 6 to 10 dB was observed. In the spectrum of the FG signal for VVM the 58 Hz component was as strong as the 116 Hz component. This result was expected since the vocal tract transfer function increases with frequency up to the first formant. Still, the 58 Hz tone was perceptually quite salient (Moore, 1997).

Overtone singing with VVM source

Figure 7 shows a spectrogram in the range 0-4 kHz for "overtone singing" of the melody "Oh, Susanah", as performed by the same subject while keeping a constant f_0. The first formant remains at approximately 400 Hz while the melody ranges between 800-1400 Hz (partials #16-26). Two other simultaneous overtone patterns, somewhat mirroring the melody, can be seen, one between 1700-2200 Hz (partials #32-42) and the other between 3000-3500 Hz (partials #57-66). The top formant seems equivalent to a "speaker's formant" (Sundberg, 1974; Leino, 1994). The inverted melodic movement suggests that in this particular overtone technique, the tongue constriction separates the mouth into two resonant cavities. A movement of the constriction produces a volume reduction in one compartment and a volume increase in the other.

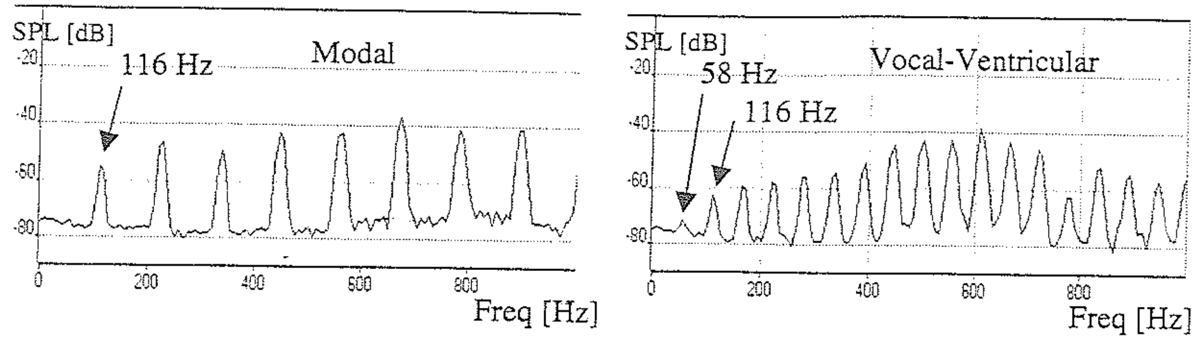

Figure 6. Detail of spectra (range 0-1000 Hz) from modal and VVM examples as in Figures 4 and 5. In the VVM spectrum, the 58 Hz "undertone" is considerably weaker than the 116 Hz partial, in general a difference of 6 to 10 dB as observed in our samples. All other harmonics of 58 Hz are present throughout the spectrum.

Fo/3 Ventricular oscillations

By laryngeal adjustment and by choosing a higher f_0, the subject was able to flip to a pitch that was a duodecime (19 semitones) below the starting pitch. Figure 8 shows the audio signal, its spectrum and the EGG recording for a typical example produced on the vowel /□/ at an f_0 of 176 Hz (F3). The audio and EGG signals are complex but periodic. In each cycle, the EGG waveform indicates short instances of maximal glottal opening interspersed with two apparently incomplete openings, i.e., reflecting an intermediate degree of glottal tissue contact. The spectrum contains multiples of 59 Hz, approximately, starting from 176 Hz. Thus, the 59 Hz and 116 Hz components are missing. Perceptually, this signal sounds as having the same pitch as the one shown in Figure 5, Bb_1 (59 Hz), although with a different tone quality. Interestingly, the sub-cycles (the peaks in the EGG waveform) indicate a quasi-periodic jitter. Unfortunately, the subject could not produce this type of phonation with the rigid telescope, so no high-speed imaging could be made.

Pulse register

Examples of this type of phonation showed a high degree of regularity. Phonation started in a steady modal tone which then suddenly switched into $f_0/2$ (Figure 9) or $f_0/3$ (Figure 10). For the example shown in Figure 9, the initial f_0 was a $C\#_3$ (approx. 138 Hz) and the period-

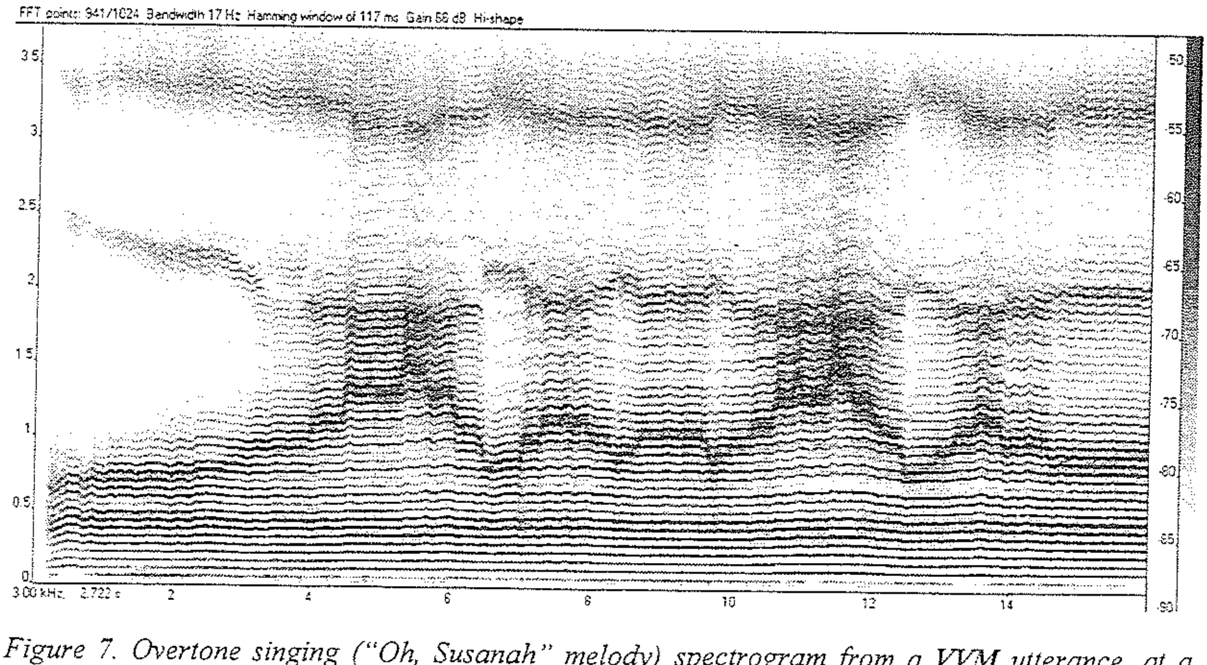

Figure 7. Overtone singing ("Oh, Susanah" melody) spectrogram from a VVM utterance, at a rather fixed ventricular frequency, approx. 53 Hz, produced by the same subject.

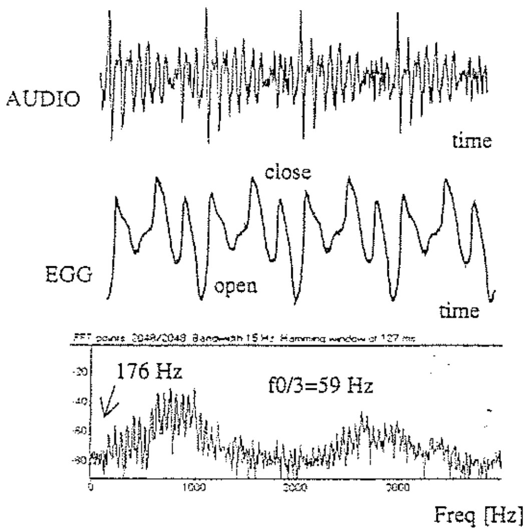

Figure 8. VVM $f_0/3$ example: waveform, EGG and power spectrum. Vocal folds at 176 Hz; ventricular folds presumably at 59 Hz.

doubling yielded an audio signal containing multiples of 69 Hz, starting from 138 Hz. For the example in Figure 10, the voice started at 207 Hz (G#$_3$) and again the transition resulted in 69 Hz modulating frequency ($f_0/3$), the spectrum containing multiples of 69 Hz that started from 138 Hz. Both these examples were perceived as having the same pitch, but different sound qualities. The highest sound power that could be produced was considerably lower in the pulse register than in the VVM. The EGG shows striking differences if compared to Figures 5 and 8, although still revealing the origin from a multiplication of the period time by 2 or 3.

Oesophageal pressures

While repeatedly alternating between modal and $f_0/2$ VVM, starting on Bb$_2$ (116 Hz) in modal register, the subject tried to keep a constant SPL of 79 dBA @ 1 m (SD = 1.0) for the modal tone. The subject also tried to constantly produce a loud VVM tone, resulting in an SPL of about 84 dBA. For the phonations in modal register, the average oesophageal pressure was 18.0 cm H_2O (SD = 0.6) while for the VVM, it was 21.9 cm H_2O (SD =1.0) These averages were calculated from four consecutive transitions.

These pressures are rather high as compared to the ones used in normal phonation or even in singing (Van den Berg, 1956; Cleveland & Sundberg, 1985). Apparently, an initially pressed voice was used in the modal register, which seemed necessary for the subject in order to facilitate well controlled and sudden transitions.

Discussion

The data presented in this report refers to a single healthy subject using three different phonation modes, modal, vocal-ventricular and pulse register. Obviously, inter-individual and even intra-individual variations can be expected. Yet, it seems likely that similar observations would emerge from studies of other subjects

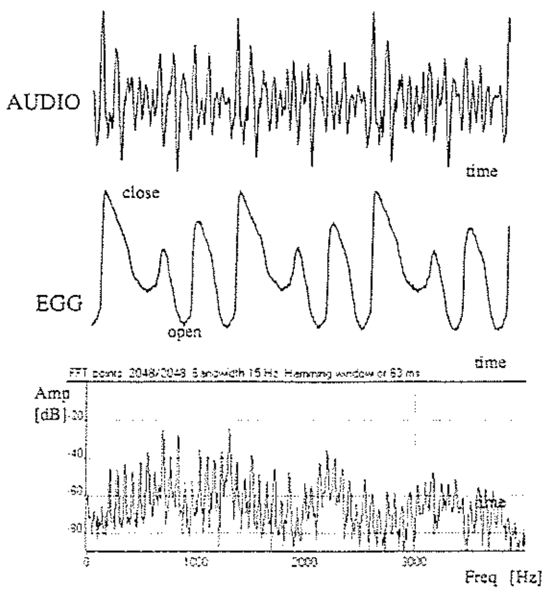

Figure 9. $f_0/2$ Pulse register (Strohbass), starting at C#$_3$ (138 Hz) and switched to C#$_2$ (69 Hz). Notice the difference in EGG pattern respect to Figure 5.

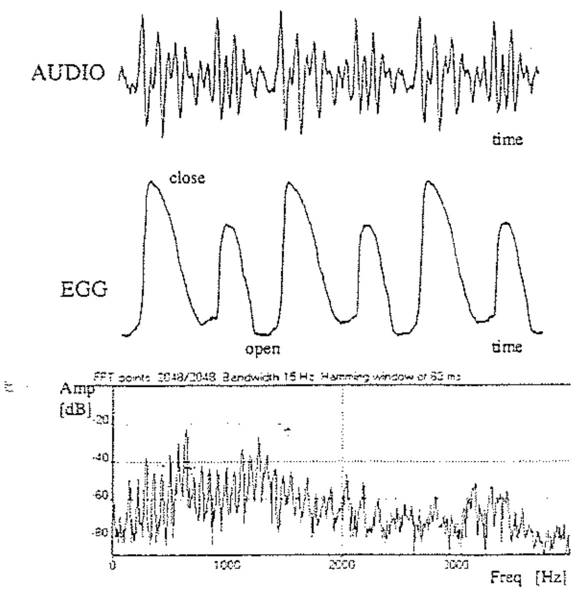

Figure 10. $f_0/3$ Pulse Register (Strohbass), starting at G#$_3$ (207 Hz) and switched to C#$_2$ (69 Hz). Compare the EGG signal with Figure 8.

producing the same types of phonation. On the other hand, ventricular dysphonia may be characterised by other laryngeal configurations, since the organic and functional prerequisites may differ from a normal subject. The advantage of focusing on these particular steady and self-sustained effects is that they provide a more robust platform to examine mechanisms which otherwise would present higher instability and transient states.

The subject is an adult male, age 36. It is an interesting issue whether the ability of producing VVM is related to gender and/or age. It has been reported that a female professional singer was capable of such phonation at 130/65 Hz (Barnett, 1977). The observed lack of female singers performing such effects in ethnical music recordings may be simply due to tradition.

Regarding vocal health, the subject never experienced any voice problems even after extensive use of the VVM. Still, attention is of course recommended to those who use all kinds of extended vocal techniques.

Period multiplication

Chaos theory has been found a useful tool in describing phenomena associated with complex vocal fold vibration (e.g., Berry et al., 1994; Davis & Fletcher, 1996). Bifurcations and chaos have been identified in the cries of newborns (Herzel & Reuter, 1996) and in asymmetric vocal fold oscillations (Steinecke & Herzel, 1995). It would be worthwhile in future research to apply chaos theory also to VVM.

Period multiplication, e.g., $f_0/2$ or $f_0/3$, occurs also in pulse register phonation (Barnett, 1977; Titze, 1994). This register has been studied and described by several authors (e.g. Hollien, 1974; Titze, 1994; Keidar, 1986; Blomgren & Chen, 1998). However, accurate measurements of the degree of vocal fold adduction, stiffness of the vocal fold margins and their precise movement are still required for a complete understanding. Investigation of the role of the ventricular folds in the pulse register mechanism has been proposed by Blomgren & Chen (1998).

The EGG waveform patterns were found to clearly differ between VVM and pulse register. In VVM the vocal and ventricular folds apparently closed in opposite phase, while in the pulse register EGG no indication of such opposite phase vibration could be observed. It would be interesting to apply physical modelling to both VVM and pulse register.

Vocal "growl" seems to be a related vocal effect, that is sometimes used by jazz and pop musical singers, e.g., Louis Armstrong. Phonation is usually breathy, complemented by a constriction of pharyngeal/laryngeal structures, as revealed by video-stroboscopy (Thalén M, personal communication). This supraglottal valve mechanism produces a modulation of the airflow, and hence of the sound quality, adding irregular or regular oscillations at low rates to the sound. In some cases, the "growl" vibrations may present a phase lock to the vocal folds in the $f_0/2$ or $f_0/3$ modes. Spectral analysis of some of Armstrong's recordings did not show similarity with our VVM examples, although some subharmonics were observed.

Terminology

A vocal tradition in different regions in Central Asia such as Tibet, Tuva, Mongolia and Ladakh have been reported to include very low pitched drones, often complemented by shifting, salient high pitched partials (Zemp, 1996; Bretèque, 1988; Bloothhooft, 1992). The same or similar effects have been referred to by many other terms, including overtone or diphonic singing (Smith et al., 1967), or "dyplophony" (Dmitriev, 1983), "throat-singing", etc. A curious and somewhat humorous variant is the voice of the cartoon character *Popeye*, who had a very low-pitched and harsh voice (down to Bb_1, 58 Hz, during singing), possibly produced by VVM. A thorough analysis of the kind applied in the present investigation would be needed for an exhaustive description of the phonatory characteristics of these tone production modes. A coherent terminology should be based on the results of such analyses.

Towards a physical model for VVM

The vibrational mechanism of the vocal folds can be represented by models, such as the classic two-mass model (Ishizaka & Flanagan, 1972), the partial differential equations models (e.g. Titze & Talkin, 1979), the multiple coupled oscillators model (Story & Titze, 1994) and the simplified two-mass model (Steinecke & Herzel, 1995). All these models imply that adjacent oscillators in the vocal folds are mechanically linked by springs, as in the lower part of Figure 11. In the case of VVM, it seems inadequate to represent the anatomically distant oscillators as adjacent, as they are separated by the laryngeal ventricle. The comparatively flaccid structure of the ventricular folds and the motional configuration indicated by the high-speed glottograms, suggest that they close because of the negative pressure generated during the open phase of the vocal folds. This force is probably strong enough to produce the collision in a "superior glottis" that is skilfully shaped by the subject. Therefore, the mastering of the VVM technique may involve the simultaneous control of adduction and positioning both at the glottal and the ventricular levels, thus generating self-sustained oscillations.

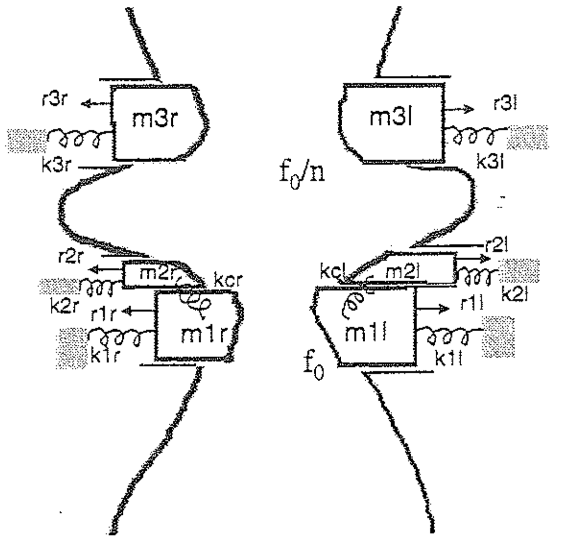

Figure 11. A proposed model for the VVM mechanism. The lower part, referring to the vocal folds, is a simplified two-mass model (Steinecke and Herzel, 1995), while the upper part consists of a pair of single masses excited by the aerodynamical forces. The relevant elements are the masses (m), springs (stiffness k) and dampers (resistance r). Indexes r and l refer to right and left sides, respectively.

A simplification of the two-mass model is that it considers the larynx as having a bidimensional or a prismatic geometry. A real larynx has a varying mass distribution along the antero-posterior axis and asymmetries with respect to the sagittal plane (Steinecke & Herzel, 1995). These variations may be responsible for several other phonation modes that will be hard to predict by means of a simple glottal model.

In Figure 11, a model for the VVM mechanism is proposed. The lower part, the glottal system, consists of the simplified two-mass model (Steinecke & Herzel, 1995). Each component contains a mass element (m), a spring with stiffness (k) and a damper (r) absorbing mechanical energy. The two pairs of glottal masses m_{1l}-m_{1r} and m_{2l}-m_{2r} are also interconnected by a spring (kc). For simulation purposes, the coefficients m, k, and r, together with the subglottal static pressure and dimensional parameters, may be used to build a system of differential equations that can be solved and predict positions and velocities for the mass pairs (Tigges et al., 1997). The ventricular system, upper part of Figure 11, contains a couple of mass-spring elements. As seen, no mechanical links have been assumed between the glottal and ventricular systems, which, however, are aerodynamically connected. This model, which needs future implementation, would be able to oscillate within a range of frequencies, but should tend to establish a harmonic relationship with f_0, thus producing a stable, self-sustained mode. Thereby, the ventricular system will achieve the "undertone".

The natural frequency of a mass element may be roughly expressed by:

$$f_0 \approx \frac{1}{2\pi}\sqrt{\frac{k}{m}} \qquad \text{Equation 1}$$

where k is the stiffness of the spring and m the oscillating mass (Figure 11). Eq. 1 does not contain any damping term and is valid for small amplitudes. For instance, during a VVM at $f_0/2$, the ventricular part can be postulated to possess a k/m ratio of approximately ¼ of the effective glottal ratio. Similarly, a VVM at $f_0/3$, the expected ventricular k/m should be approximately 1/9 of its glottal k/m equivalent.

Acoustically, these alternated closures of the vocal and ventricular folds produce different effects.

One is the partial damping of every second (or third) vocal pulse. This factor is responsible for the period multiplication and for the amplitude modulation at f_0/n.

The other effect is the increase of the effective close phase, yielding a longer period for reflection of the standing waves in the vocal tract between cycles. This results in a narrowing of the formant bandwidths, which should add to the resulting sound level and reduce the decay of the spectrum envelope.

The presence of two cascaded oscillating valves differing in phase should increase the flow resistance of the system, thus requiring higher subglottal pressures. This is in accordance with the oesophageal pressure measurements. It may be speculated that VVM phonation is facilitated at high lung volumes, since these are associated with high relaxation pressures (Agostoni & Mead, 1965); Tibetan chant, presumably produced in VVM phonation, is used for meditation.

A VVM at $f_0/1$, i.e., with the ventricular folds oscillating at the same frequency as the vocal folds, may be possible. Then, every single glottal pulse would be damped in a similar way, probably generating a highly muted sound.

Another possibility is that the ventricular folds vibrate at a multiple of the vocal frequency, i.e. n x f_0. In this case, the ventricular folds would serve as the main oscillator. These speculations call for additional experimental data and further investigation.

Vocal registers

The above findings raise the question whether or not VVM should be regarded as a special vocal register. Unfortunately, there is no general agreement on vocal register terminology. According to Hollien (1974), there are no more than three registers (a) loft, also referred to as falsetto; (b) modal or chest register; and (c) pulse register, also called vocal fry, creaky voice or, in singing, strohbass. From our observations, the VVM phonation does not seem similar to any of the above registers. Thus, it appears reasonable to accept it as a register of its own.

Musical implications

Some contemporary music makes use of special vocal techniques adding new sonorities to the available resources. We have demonstrated that the VVM is generated by a specific mechanism and can be produced at various degrees of loudness over an f_0 range of about one octave, starting from 49 Hz, approximately. These characteristics may seem attractive to some contemporary composers.

The VVM should also be applicable to the so-called "hum and play" technique, where wind instrumentalists play the instrument while vocalising. This technique is used in several wind instruments, such as brass, flute, saxophone and clarinet (e.g. Berio's *Sequenza V* for solo trombone, William Smith's *Variants* for solo clarinet). By changing the fingerings and lengths of the instruments, several kinds of "filters" may be defined, resulting in a gamut of new sonorities. The VVM phonation yields powerful and spectrally rich sounds, and may provide a useful extension of the present timbral possibilities.

Conclusions

This exploratory study described a particular phonatory mode produced by a healthy trained subject with simultaneous oscillations of the vocal folds and the ventricular folds.

The frequency ratio between vocal and ventricular folds was 2:1 or, in some cases, 3:1. The closure of the ventricular folds occur during the open phase of the vocal folds, thus damping every second and/or third glottal pulse. The driving of the ventricular folds seems to be a negative pressure generated by the airstream contained in the glottal pulse. A tentative model was presented to describe the mechanism. According to an expert listener, the vocal effect was similar to that produced in some monasteries of Tibet.

The subglottal pressures required were consistently higher than those used in modal and pulse register, at least in the subject used in this experiment. The pitch range approached one octave starting at 49 Hz (G_1), approximately. The SPL @ 0.3 m could be varied within 64 and 88 dBA for the 58Hz (Bb_1) tone.

This technique may appear attractive to contemporary music, and can also be employed in wind instruments playing, producing a family of new sounds.

Acknowledgements

The kind and expert assistance of Stellan Hertegård, from Karolinska Hospital, since the first stages of this investigation, including preliminary laryngeal examination and video-stroboscopic recording, is gratefully acknowledged. The long-term technical discussion with Svante Granqvist on acoustical aspects of VVM was helpful and stimulating. Lobsang Tenzin Kyaga kindly judged live demonstrations of various types of vocal sounds. Per-Åke Lindestad skillfully applied the high-speed imaging. Joakim Gustafson assisted in the digital conversion of the video files and Jenny Iwarsson assisted in the inverse-filtering, and Maria Södersten and Hans Larsson from Huddinge University Hospital, Karolinska Institute, offered technical assistance during the high-speed imaging recording. The high-speed imaging was carried out at the Karolinska Institute, Department of logopedics and Phoniatrics, supported by a grant from the National Medical Research Council, Sweden. The author profited from valuable discussions with Gunnar Fant, Johan Liljencrants, Anders Askenfelt and Jocy de Oliveira. This research was supported by a scholarship from the Brazilian Ministry of Education, through Capes Foundation, and by the Rio de Janeiro Federal University, UFRJ, School of Music. This manuscript is dedicated to Bonnie Barnet.

References

Agostoni E & Mead J (1965). Statics of the respiratory system. In: *Handbook of Physiology, Respiration*, Am Phys Soc, Washington, chapter 52: 1875-1886.

Barnett BM (1977). Aspects of vocal multiphonics, *Interface* 6: 117-149.

Berry DA, Herzel H, Titze I & Krischer K (1994). Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions, *J Acoust Soc Am* 95: 3595-3604.

Blomgren M & Chen Y (1998). Acoustic, aerodynamic, physiologic and perceptual properties of modal and vocal fry registers, *J Acoust Soc Am* 103/5: 2649-2658.

Bloothooft G, Bringmann E, van Cappellen M, van Luipen JB & Thomassen KP (1992). Acoustics and perception of overtone singing, *J Acoust Soc Am* 92/4: 1827-1836.

Bretèque BA de la (1988). Le chant diphonique, *Revue de Laryngologie* 109/4: 387-388.

Cleveland T & Sundberg J (1985). Acoustic analysis of three male voices of different quality, *SMAC83 Conference Proceedings*, Stockholm, 1: 143-156.

Davis P & Fletcher N (editors) (1996). *Vocal fold physiology-controlling complexity and chaos*, a collection of invited papers presented at the 9th Vocal Fold Physiology Symposium, Australia-1995, San Diego: Singular Publishing Group.

Dmitriev LB, Chernov BP & Maslov VT (1983). Functioning of the voice mechanism in double-voice Touvinian singing, *Folia Phoniat* 35: 193-197.

Eysholdt U, Pröschel U & Tigges M (1994). Direct evaluation of high-speed recordings of vocal fold vibrations. *Proceedings of the 3rd International Symposium on Phonosurgery*, Kyoto, Japan.

Fabre P (1957). Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de fréquence; premiers résultats. *Bull Acad Nat Méd*, 141: 66-69.

Fant G (1960). *Theory of Speech Production*, The Hague: Mouton.

Fant G (1997). The voice source in connected speech, *Speech Communication* 22: 125-139.

Fant G, Liljencrants J & Lin Q (1985). A four-parameter model of glottal flow. *STL-QPSR, KTH* 4/1985: 1-13.

Freud ED (1962). Functions and dysfunctions of the ventricular folds, *J Speech Hear Dis* 27/4: 334-340.

Fry DL, Stead WW, Ebert RV, Lubin RI & Wells HS (1952). The measurement of intraesophageal pressure and its relationship to intrathoracic pressure. *J Lab Clin Med* 40: 664-673.

Herzel H & Reuter R (1996). Biphonation in voice signals, non-linear chaotic and advanced signals processing methods for engineers and scientists. Woodbury: AIP Press.

Hollien H (1974). On vocal registers, *Journal of Phonetics* 2: 125-143.

Leino T (1994). Long-term average spectrum study on speaking voice quality in male actors, *Proceedings of the Stockholm Music Acoustics Conference SMAC 93*, Royal Swedish Academy of Music, No 79: 206-210

Lindestad P-Å, Södersten M, Merker B & Granqvist S (1998). Mongolian "throat singing"- acoustical and vibratory characteristics, *Proc. 24th Conference of International Association of Logopedics and Phoniatrics*, Amsterdam (forthcoming).

Moore BCJ (1997). *An introduction to the psychology of hearing*, San Diego: Academic Press; 188-189.

Nasri S, Jasleen J, Gerratt BR, Sercarz JA, Wenokur R & Berke GS (1996). Ventricular dysphonia: a case of false vocal fold mucosal travelling wave. *Am J Otolaryngol* 17/6: 427-431.

Orlikoff RF (1998). The uses and abuses of electroglottography, *Phonoscope 1/1: 37-53.*

Pressman JJ & Kelemen G (1955). Physiology of the Larynx, *Physiol Rev* 35: 506-554.

Rohen JW & Yokochi C (1993). *Color Atlas of Anatomy*, 3rd ed., New York: Igaku-Shoin Med Pub.

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

Smith H, Stevens KN & Tomlinson RS (1967). On an unusual mode of chanting by certain Tibetan Lamas, *J Acoust Soc Am* 41/5: 1262-1264.

Steinecke I & Herzel H (1995). Bifurcations in an asymmetrical vocal-fold model, *J Acoust Soc Am* 97/3: 1874-1884.

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

Tigges M, Mergell P, Herzel H, Wittenberg T & Eysholdt U (1997). Observation and modelling of glottal biphonation, *Acustica/Acta acustica*, 83: 707-714.

Titze IR (1981). Biomechanics and disturbed-mass models of the vocal fold vibration. In: Stevens K & Hirano M, eds, *Vocal fold physiology, Proceedings of the vocal fold physiology conference*, University of Tokyo, Japan; 245-270.

Titze IR (1994). *Principles of voice production*, N.J: Prentice Hall.

Titze IR & Strong WJ (1975). Normal modes in vocal chord tissues. *J Acoust Soc Am* 57: 736-744.

Van den Berg JW (1956). Direct and indirect determination of the mean subglottic pressure, *Folia phoniatr* 8: 1-24.

Zemp H (editor) (1996). *Voices of the world - an anthology of vocal expression*, Paris: UMR 9957-Musée de l'Homme.