SINGLE LINE SCANNING OF VOCAL FOLDS AS FEEDBACK IN SINGING: THE 'MESSA DI VOCE' EXERCISE

P.H. Dejonckere [1], J. Lebacq [2], L. Bocchi[3], C. Manfredi [3]

[1] University of Leuven, Neurosciences, Exp. ORL, Belgium ; Federal Institute of Occupational Diseases, Av. De l'Astronomie, 1, B-1210 Brussels Belgium.
[2] Institute of Neuroscience CEMO, Université Catholique de Louvain, Avenue Hippocrate 55 bte B1.55.12 B-1200 Brussels, Belgium.
[3] Department of Information Engineering, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy.

Abstract: **This article describes a novel application of the 'single line scanning' of the vocal fold vibrations (kymography) in singing pedagogy, particularly in a specific technical voice exercise: the 'messa di voce'. It aims at giving the singer relevant and valid short-term feedback. An user-friendly automatic analysis program makes possible a precise, immediate quantification of the essential physiological parameters characterizing the changes in glottal impedance, concomitant with the progressive increase and decrease of the lung pressure. The data provided by the program show a strong correlation with the handmade measurements.**
Additional measurements as subglottic pressure and flow glottography by inverse filtering can be meaningfully correlated with the data obtained from the kymographic images.

Keywords: **Messa di voce, single line scan, VKG, inverse filtering, subglottic pressure, singing.**

I. INTRODUCTION

This article describes a novel and original application of the 'single line scanning' of the vocal fold vibrations (kymography) in singing pedagogy. Short-term feedback for relevant physiological parameters of voice production may be very useful for the singer, particularly in acquiring specific technical skills based on motor control, e.g. for producing a 'messa di voce'.

Single line scanning of vocal fold vibrations (kymography, or videokymography: VKG) [1] is an imaging method based on a special digital camera, which can operate in two different modes: standard and high-speed. In the standard mode, the camera provides images displaying the whole vocal folds at standard video frame rate (30/25 frames/s, with 720x486/768x576 pixel resolution). In the high-speed mode, the video camera delivers images from a single line selected from the whole image, at the speed of approximately 7875/7812.5 line-images/s and 720x1/768x1 pixels resolution. The selected line is usually at the level of the midportion of the vibrating folds. The technique allows a clear visualisation of some essential physiological parameters of vocal fold vibration: period, duration of opening, closing and closed phases, amplitude of the vibration and right-left symmetry (Fig.1).

Kymography has been applied successfully to voice pathology [e.g. 2 – 5]. However, another potential field of applications of the technique is singing voice pedagogy, as it provides real time visual feedback for the subject, and allows short term inspection and analysis of the recorded sequence. The method becomes particularly interesting with the input of automatic quantitative analysis of the above-mentioned parameters, as it has already been achieved for pathology, however focusing on asymmetries and irregularities [2,5].

The 'messa di voce' is a gradual crescendo and decrescendo on a sustained (sung) note , and is known as one of the most difficult exercises for singers. This exercise is frequently associated with the 'candle test', a centuries old method used by singers to evaluate airflow. It simply consists of singing a vowel with a candle flame placed about five inches from the mouth. Depending on the exhaled airflow, the flame will waver very little (or not at all) or flutter wildly [6]. During a 'messa di voce' by a trained singer, the flow increase at the *ff* must remain minimal.

'Messa di voce': the underlying physiology

From a physiological point of view, a progressive increase in voice intensity – thus primarily in subglottic pressure - needs several critical requirements [7 ; 8]: (1) a constant adjustment of tension in the intrinsic laryngeal muscles, particularly the m. vocalis and the m. cricothyroideus, in order to keep the fundamental frequency (Fo) constant and to compensate for the enhanced passive strain in the folds, induced by the averaged higher vibration amplitude; (2) a specific mechanical regulation at the transition from a 'one-mass-model' vibration in *pp* (falsetto-like) to a 'two-mass-

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 8 th international workshop : December 16-18, 2013*
ISBN 978-88-6655-469-1 (print) ISBN 978-88-6655-470-7 (online) © 2013 Firenze University Press

model' (usual modal register), particularly at the early beginning of the crescendo; (3) a permanent control of the voice quality in order to avoid audible noise appearance due to increase of the transglottic airflow, concomitantly with the increase of the lung pressure. Actually, the performer must succeed in obtaining at each intensity level an exact balance between expiratory pressure and glottal impedance. Ideally, it results in only a slight increase in transglottic flow [9].

II. Material & Methods

(1) Kymography system: The single line scanning system used in these experiments is comprised of a Lambert CCD-Kymocam with technical characteristics corresponding to the above-reported description, a rigid 90° Wolf laryngeal telescope, a JVC-magnetoscope and a monitor. The telescope has a magnifying facility, with narrow depth of field and critical sharpness adjustment.

Vocal material: A trained vocalist (baritone), was asked to produce series of 'messa di voce' utterances on different pitches, avoiding to elicit vibrato. Vocal fold vibrations were recorded using the VKG system and the subject could have a real time visual control on the screen. After some preliminary trials, the subject became able to handle the scope himself and to find the optimal placement, combining comfort and quality of vocal fold image. During four different sessions 62 recordings were achieved. The sound was also recorded by a Sennheiser MD 4210 microphone at 10 cm of the lips, for analysis of SPL and Fo (using PRAAT 5.3.10, 2012 by P. Boersma & D. Weenink: www.praat.org.). One of the recordings, considered as representative and providing a complete visualization of the endolarynx was selected. For demonstration 40 single line scans, each showing two vibratory cycles, were taken at 125 ms intervals in order to cover the complete utterance (5 s). Fig. 1 shows two characteristic kymograms, respectively at the early beginning of the 'messa di voce' and at approximately the maximal SPL of the 'messa di voce' (the closed phase near half of the period).

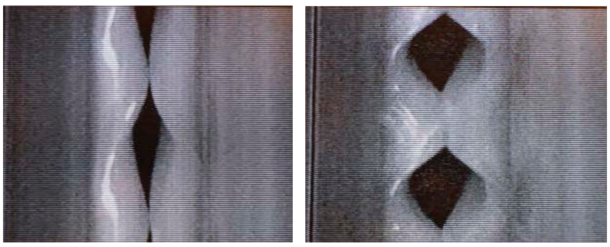

Fig. 1: Two examples of single line scans during a 'messa di voce'. Left at the beginning (*pp*), right at *ff*.

These images were enlarged and printed for manual measurements of the following parameters: right and left period, right and left amplitude, open and closed phase, right and left opening time. All measurements were made manually, independently by two observers, and averaged. The closed quotient is the quotient 'duration of closed phase / duration of cycle'. An adequate calibration was achieved with respect to time and distance. For distance, scale paper was filmed without changing the focusing.

(2) Program for automatic analysis: First attempts to automatically analyze single line scans of vocal fold vibration were achieved by Qiu & al. [10]. The approach used is based on monodimensional active contours, where each line has its own energy to be minimized separately from other lines. With the VKG-Analyser used in the current experiments, a planar active contour is applied, i.e., the set of all points (two for each VKG line) is considered as a pair of lines (corresponding to left and right vocal folds) that contain a surface. During subsequent iterations, the contour is modified in order to adapt to the shape given by the dark pixels of the image. Planar snakes allow the algorithm to find a global minimum of the energy, thus performing a global optimisation instead of a series of local ones. Moreover, in the VKG-Analyser, specific parameters are evaluated for the case of incomplete vocal fold closure. A detailed description of the VKG-Analyser has been provided in a recent publication [5]. Each image in the sequence can be processed using a digital image processing algorithm developed and optimized for the analysis of VKG recordings. It performs intensity adjustment, noise removal and robust techniques for vocal fold edge detection to avoid fluctuations of the grey levels in regions at a distance from the vocal folds.

(3) The vocal folds contour detection algorithm consists of two main steps: the first one defines an initial contour of the glottal area opening, using an adaptive threshold. A refining iterative procedure, based on active contours, is applied to the region, giving the final segmentation. The control parameters which drive both steps are determined automatically by the program. However, the user can manually adjust some of the controls for improving segmentation using a set of controls present in the user interface. The software allows to individually select the desired frame(s) to be processed from the video recording. Once the final contour has been obtained, the parameters of interest are evaluated. The software is designed so as to give a value of each parameter for each video frame by averaging the parameters over all the vibration periods which can be observed on the frame. This reduces the variability of the results smoothing out noise and eases the management of data giving a fixed number of values for a given video sequence,

independently from the acquisition. The program also allows storage and retrieval of subject's data, display of tracked parameters, results and statistics.

(4) Additional measurements

For the purpose of this study, the vocalist – once trained in optimizing his 'messa di voce' - repeated the exercise for separate indirect measurement of subglottic pressure and flow-glottography. For indirect measures of subglottic pressure, the short flow interruption method was used [11]. The glottal volume velocity waveform was recorded with a Rothenberg mask and the MSIF2 inverse filtering system of Glottal Enterprises, Inc.

III. RESULTS

Fig. 2 shows the basic acoustic characteristics of a typical 'messa di voce'. Amplitude increases and decreases (~ 35 dB at 10 cm) while Fo remains stable (165 Hz). Duration of the sequence is about 5 s.

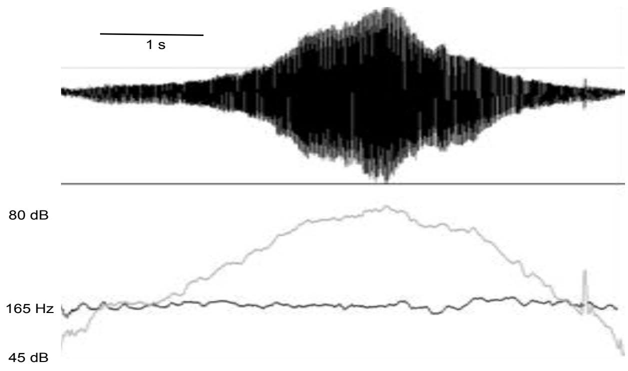

Fig. 2: messa di voce: microphone signal, Fo and intensity (duration about 5s).

Changes in % over time in different physiological parameters can be displayed: amplitude of vibration (Right + Left), duration of the closed phase and closed/open quotient (quotient of duration of the closed phase / duration of open phase), opening and closing speed etc. Fig. 3 shows as an example the evolution of the closed/open quotient, with as well the measures made by hand as those made by the computer program.
The lowest value for each parameter is set as 100%, in order to illustrate the relative changes. The smoothing curves (least squares fit) are shown as solid lines. The correlation coefficient (manual / automatic) is 0.88. Similarly but for amplitude measurements, the correlation coefficient is 0.84 (Fig. 4). Results of the additional instrumental measurements are shown in Figs. 4 and 5: estimates for subglottic pressure vary between about 3 and 20 hPa. There is a satisfactory control of the air flow around approximately 200 – 250 ml./s.

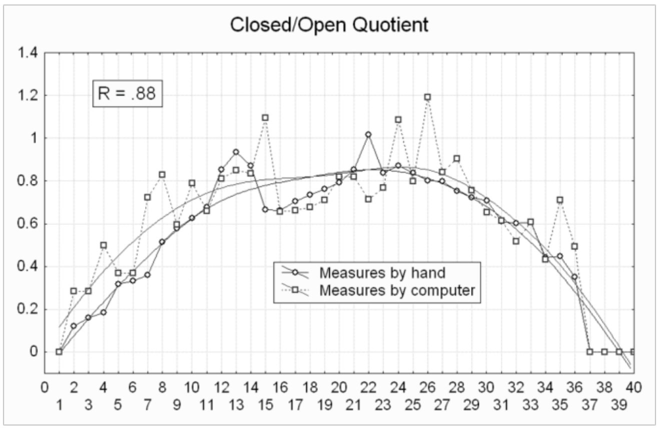

Fig. 3: Evolution over time (5s) of the closed/open quotient. Comparison of measures by hand and by computer

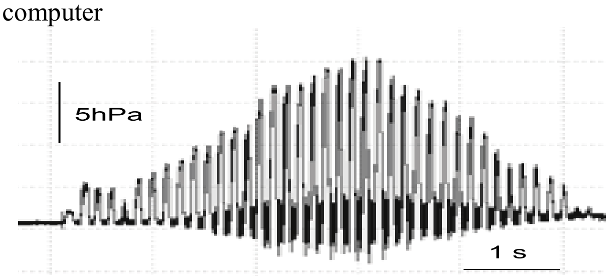

Fig. 4: Estimate of subglottic pressure by measurement of the intraoral pressure (flow interruptions method).

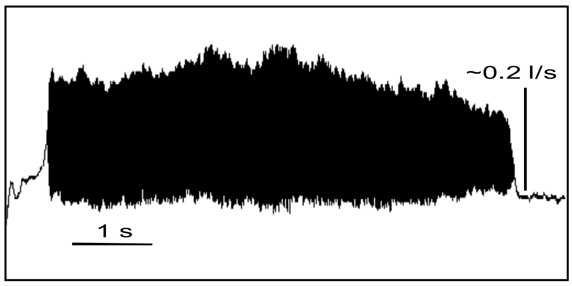

Fig. 5: Glottal volume velocity waveform measured with Rothenberg's mask and inverse filtering.

IV. DISCUSSION

Global feedback on the acoustic phenomenon

During the execution of the 'messa di voce' the changes of intensity are, in this vocalist, clearly correlated with the changes of breath pressure [12], although changes in spectral balance (singer's formant) may contribute to perception of increased loudness. Ideally, one could expect the exercise to be performed as a symmetric triangle: a linear increase in loudness, followed by a linear decrease.

Detailed feedback on vocal fold biomechanics

The cricothyroid and thyroarytenoid muscles are primarily involved in regulating the biomechanical properties of vocal fold vibration. As the vocal intensity increases, it is expected that their tension will also be increased at a gradual pace to resist the growing breath pressure, but without compromising tonal accuracy. Videostroboscopy of the vocal fold vibration has demonstrated that the degree of glottal closure increases with intensity of phonation in normal subjects [13]. Furthermore, also in normal subjects, the glottal impedance is well reflected by the electroglottographic closed quotient, and the quotient clearly increases with voice SPL; this increase in vocal fold closure is plausibly related to an active thickening of the vocal fold edge (in modal register), and to a longer closed phase of the vibratory cycle [14]. This may account for the limited concomitant increase in transglottal flow [10]. Titze [8] has hypothesized (using computer modeling) that the maximal glottal source power should occur somewhere around a closed/open quotient of 0.5, i.e. when open and closed phases are approximately equal (as observed in the demonstrated case, see Fig. 2, right).

The need for an automatic analysis program

However, these quantitative data are of practical value only if they are available immediately after the 'messa di voce'. Therefore, in this work, a new user-friendly tool, the VKG-analyser [2;5] is used and tried out for the automatic extraction, tracking and computing of quantitative parameters from the VKG images. In the present study, comparative plots of the closed/open quotient measured manually (average of two raters) and by the automatic analysis program confirm the efficacy of the analysis program: the correlation coefficient between the two measurements is 0,88 (Fig. 3). For amplitude measurements, the correlation coefficient is 0,84.

Additional measurements

Both subglottic pressure and flow-glottography are fitted for real-time visual feedback: the direct visualization of the pressure data will help in increasing symmetry (Fig. 4) while monitoring the phonation flow plays the role of the historical candle in front of the mouth (Fig. 5).

V. Conclusion

'Single line scanning' of the vocal fold vibrations appears to be – beside its clinical usefulness - also suitable for investigating a specific technical voice exercise (and musical ornament) as the 'messa di voce'. It makes possible a precise quantification over time of the essential physiological parameters characterizing the changes of glottal impedance concomitant with the progressive increase and decrease of the lung pressure. However, introduction of a valid and user-friendly automatic analysis program of kymography-images appears indispensable for opening such new applications in the field of voice pedagogics by short-term feedback. The data provided by the automatic analysis program show a strong correlation with handmade measurements. Additional measurements (subglottic pressure and phonation flow) can meaningfully be correlated with the data obtained from the kymography-images.

References

[1] Svec J, Schutte H K. Videokymography: high-speed line scanning of vocal fold vibration. J Voice 1996; 10: 201–205.

[2] Manfredi C, Bocchi L, Bianchi S, Migali N, Cantarella G. Objective vocal fold vibration assessment from videokymographic images, Biomed. Signal Process. Control 2006; 1: 129–136.

[3] Svec J, Sram F, Schutte H K. Videokymography in voice disorders: what to look for ? Ann. Otol. Rhinol. Laryngol. 2007; 116: 172–180.

[4] Piazza C, Mangili S, Del Bon F, Gritti F, Manfredi C, Nicolai P, Peretti G. Quantitative analysis of videokymography in normal and pathological vocal folds: a preliminary study. Eur Arch Otorhinolaryngol 2012; 269: 207-212.

[5] Manfredi C, Bocchi L, Cantarella G, Peretti G. Videokymographic image processing: Objective parameters and user-friendly interface, Biomedical Signal Processing and Control 2012; 7: 192– 201.

[6] Taylor D C. A Rational Method of Voice Culture based on a Scientific Analysis of all Systems, Ancient and Modern. N.Y.,The MacMillan Co., 1922.

[7] Titze I R. Principles of Voice Production. Englewood Cliffs, NJ: Prentice Hall, 1994.

[8] Titze I R. More on Messa di voce; Journal of Singing 1996; 52: 31 – 32.

[9] Smitheran JR, Hixon TJ. A clinical method for estimating laryngeal airway resistance during vowel production. J Speech Hear Disord 1981;46:138-146.

[10] Qiu Q, Schutte H K, Gu L, Yu Q. An automatic method to quantify the vibration properties of human vocal folds via videokymography. Folia Phoniatr Logop 2003; 55: 128-136.

[11] Hertegard S, Gauffin J, Lindestadt P A . A comparison of subglottal and intraoral pressure measurements during phonation. J Voice 1995; 9: 149-155.

[12] Titze I R, Long R, Shirley G I, Stathopoulos E, Ramig L O, Carroll L M, Riley W D . Messa di voce: an investigation of the symmetry of crescendo and decrescendo in a singing exercise. J Acoust Soc Am 1999; 105: 2933 – 2940.

[13] Sodersten M, Lindestadt P A. Glottal closure and perceived breathiness during phonation in normally speaking subjects. J Speech Hear Res 1990; 33: 601-611.

[14] Dejonckere P H. Control of fundamental frequency and glottal impedance with increasing sound pressure in normal and pathological voices. Voice 1994; 3: 10-16.