EVALUATION OF THE NEO-GLOTTAL CLOSURE BASED ON THE SOURCE DESCRIPTION IN ESOPHAGEAL VOICE

Roberto Fernández-Baíllo[1], Pedro Gómez[1], Bartolomé Scola[2], Carlos Ramírez[2]

[1]GIAPSI, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n 28660 Boadilla del Monte, Madrid. e-mail: roberto@junipera.datsi.fi.upm.es

[2]Servicio ORL, Hospital Provincial Universitario Gregorio Marañón, C/ Doctor Esquerdo. Madrid.

ABSTRACT

The characteristics of esophageal voice render its study by traditional acoustic means to be limited and complicate. These limitations are even stronger when working with patients lacking minimal skills to control the required technique. Nevertheless the speech therapist needs to know the performance and mechanics developed by the patient in producing esophageal voice, as the specific techniques required in this case are not as universal and well-known as the ones for normal voicing. Each patient develops different strategies for producing esophageal voice due to the anatomical changes affecting the crico-pharyngeal sphincter (CPS) and the functional losses resulting from surgery. Therefore it is of fundamental relevance that practitioners could count on new instruments to evaluate esophageal voice quality, which on its turn could help in the enhancement of the CPS dynamics. The present work carries out a description of the voice of four patients after undergoing laryngectomy on data obtained from the study of the neo-glottal wave profile. Results obtained after analyzing the open-close phases and the tension of the muscular body on the CPS are shown.

1. INTRODUCTION

The laryngectomy is a surgical technique to remove completely the larynx of a patient as a radical treatment of larynx cancer. As a side effect the patient suffers a series of anatomical changes conveying the alteration and/or loss of certain funcions. Doubtless the strongest limitation that these patients experience is the loss of the voicing function, therefore the capability of speech communication. Therefore the process of post-surgical treatment and rehabilitation maintain the restoring of communication skills as the main goal. In these cases the speech therapist has to teach the patient a new model of voice production, which is known as erigmophonic or esophageal voice. This voicing method follows the same general principles that work in the case of laryngeal voicing, which consists in using the closure produced by a muscular sphincter to push an air column through the closure to the vocal tract, which is consistently used in producing the classical modulations in normal speech (1). The CPS is used in esophageal voice as a source of vibration (neo-glottis) as the air pushed through is previously stored in the esophageal tract. One of the limitations presented by esophageal voice is the volume of air which can be stored in the esophageal tract, to be around 50 ml under normal conditions (2), which for esophageal voicing can be augmented up to 94 ml (3). This results in a slower production of esophageal voice as compared to normal voicing, because frequent stops for air reservoir refilling are mandatory. Another important restriction in this neo-phonatory gesture resides in the radical different dynamics shown by the CPS as confronted with that of vocal folds. The CPS is a constrictor type of sphincter which attains closure by strangling the normal opening of this neo-glottis. Besides the capability of closure control and the relative muscular tenseness have to be considered. Finally it must be considered as well that the esophageal mucose is different to the vocal fold mucose. Both share a stratified epithelium on their surface albeit their physiological underlying structures are rather different.

Esophageal voice quality is strongly related with the mechanism used for voice production and its dynamics, mainly the control of the CPS tension (4). Currently the studies of the CPS dynamics are carried out by means based on radiography, manometry (5) and video-fluorescence (6). These methods present several limitations, are invasive or risky and none of them supply joint data on neo-glottal dynamics and voice quality. The present work is based on an inverse filtering method used in the estimation of the glottal source in laryngeal voicing (7) to establish a dynamic correlate of the CPS. For such it is assumed that the CPS is functionally operating in a similar way to the vocal folds and that a phenomenon of mucosal wave may be appreciated in this case as well.

2. METHODS

2.1 Sample collection

The esophageal voice of four male patients was recorded. All but one who used an injection strategy achieved esophageal phonation through the deglutition method (8) (9). A classification of the patients as a function of the acceptability of their voices from a perceptual evaluation by an expert was carried out (see Table I).
The recording protocol included the production of an utterance of the sustained vowel /a/. The phonation time for all patients was under 1.5 sec. A fragment of 0.2 sec from the primary recording was used for the neo-glottal source and the mucosal wave extraction by signal processing methods. The recording and the processing was carried out using the software GLOTTEX® (see Fig. 2).

Table I. Patient Classification.

Frame No.	Technique	Tenseness	Noise
ES-1	Deglutition	High	Low
ES-3	Deglutition	Low	High
ES-4	Deglutition	Very high	High
ES-7	Injection	High	Low

2.2 Estimating the neo-glottal source and the mucosal wave correlate (GLOTTEX®).

The main assumption used is that the model of esophageal voice production is similar to that of laryngeal voice. Therefore the esophageal voice is defined as the lip-radiated resulting wave as generated bu the CPS when excited by a column of air from the esophagus and modified spectrally by the resonant organs in the pharynx and oral and nasal cavities. The CPS sphincter is a muscular ring established between the cricoid (upper larynx limit) and the low pharynx constrictor muscle. Therefore the length of the vocal tract in the laryngectomized patient will be similar to the laryngeal voice producing model. As such the same signal processing techniques used in normal phonation for the estimation and removal of the vocal tract used in normal voicing will also be of application in esophageal voicing (7), as the vocal tract will not have undergone important transformations after surgery. The result obtained after the vocal tract transfer function removal from voice will be a neo-glottal signal (source) resulting from the CPS vibration, although the dynamics of closure of the CPS (being an orbicular muscle) will be rather different as compared to that of the vocal folds. Once the closure of the air passage is attained the dynamic behaviour of both systems will be similar. In both cases a muscular body and a cover layer may be identified. The vibration of both the vocal folds and the CPS will be produced by the pass of an air column displacing the upper and lower components of the cover with a different phase shift. Therefore the same k-mass models used in the study of the dynamics of the vocal folds (7) can be applied to approach the study of the CPS dynamics. The body-cover model allows the estimation of the tensions present in each dynamic element during the production of the voice. Therefore high tensions on the cover will diminish the amount of the mucosal wave component and high tensions on the body are related with a high pressure during closure. The signal processing method is embedded in the software GLOTTEX, which generates a correlate of the glottal source (neo-glottal in the case of esophageal voice) on which two different components can be appreaicated (see Fig. 1): a one-phonation-cycle

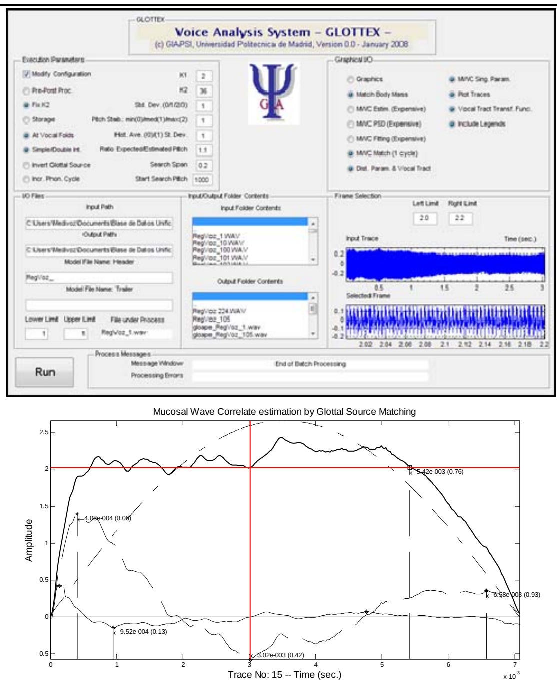

Figure 1. *The upper template shows the Graphical User Interface of the software used for recording and signal processing. The lower template is the glottal source for non-pathologic laryngeal male voice. The reference points for the study of the open and closed phases are drawn on the picture. It may be appreciated that the appearance of the glottal source profile approaches that of the well-known L-F model (11).*

signal (interval between two consecutive neo-glottal closures, including closure and opening phases) which is referred by Titze (13) as the Average Acoustic Wave, and a component which preserves the high frequency contents of the (neo) glottal source which is referred to as the dynamic cover component or conversely the mucosal wave correlate. In Fig.1 an example of laryngeal glottal source is shown, whereas several realizations of esophageal neo-glottal source are given in Fig. 2.
The software used allows the extraction of a series of singular points on the glottal source profile, which are useful in detecting the close and open phases in reference with the classical L-F model (Fig. 1). The application allows estimating a series of parameters related with the biomechanics of k-mass models of vocal fold system (masses, tensions and losses) both on the fold body and cover (7) (Fig. 2).

3. RESULTS AND DISCUSSION

The profile of the glottal source for non-pathologic laryngeal voice is characterized by its proximity to the ideal L-F model (12). The maximum amplitude value on this profile corresponds with the open phase (Pho), whereas the amplitude during the close phase (Phc) is uniform and smaller than that during Pho. This condition was also observed in all the recordings of esophageal voice analyzed (see Fig. 2), independently of the voicing technique used by each patient. In all cases Pho presents larger amplitudes than those found for Pho, although differences in amplitude and duration can be found among the different cases. The variability obtained in Pho is directly related with the kind of closure produced. Those closures which seem to be more absolute and plosives as in Esophagea-1 produce a longer and more energetic Pho, whereas a more relaxed sustained closure results in a shorter and smaller Pho as in Esophagea-2. When the profiles of the neo-glottal source shown in Fig. 2 are related with the technique used in the production of esophageal voice (Table I) a difference in the duration of the Phc among the patient using the injection technique and those using deglutition could be observed. These results are justifiable on the mechanism behind each technique, as the deglutition method consists in swallowing air and when its insertion in the esophagus is perceived the process of ejection may start while the CPS is still closed. This is an explosive technique which a skilled erigmophonic speaker uses during the early phases of training, to be later abandoned for its slowness and the low efficient use of air in each emission. The patients producing Esophageal-1 and Esophageal-4 are good examples of what has been discussed, as in both cases a plosive Pho can be observed. The injection method is based on the use of the articulatory tension to sip air into the esophagus. This method is more difficult but once controlled it results in a more fluid production of speech. In this case the CPS closure is not very tense but more sustained in time. The results can be seen in Esophageal-2 which represents a good example of what has been exposed. The noise component in the laryngectomized voicing is directly related with the closure efficiency and with the technique and tension used as well. The deglutition method and large tension usually produce noise at the level of the osteoma, which can be found in the produced voicing. In patients producing Esophageal-3 and Esophageal-4 the initial perceptual analysis detected high levels of noise, as in both cases the Phc is less complete, and present energy losses which convey the production of noise.

Many authors have referred to the importance of the CPS pressure in the final quality of esophageal voicing (4). These authors address to a certain critical pressure for the production of esophageal voice. Values over this critical point around 17 mmHg accordingly to Frint et al. (14) or 13 mmHg following Winans et al. (15) will difficult or render the production of esophageal voice impossible. On the other side Aguiar-Ricz et al. (16) conduct a study on laryngectomized patients and their results do not show any kind of relationship between the CPS pressure and the production of esophageal voice. In the present study the direct evaluation of the CPS could not be carried out, but the tension of the mucosal cover and muscular body of the sphincter can be indirectly estimated by inverse methods. The CPS as already explained works as an orbicular muscle, therefore it can be assumed that the higher the contraction degree the larger will be its tenseness and the higher the pressure at the level of the CPS. In Fig. 2 values of the tenseness correlate on the muscular body (BSC) of the CPS are shown for all the cases recorded. The results of this work coincide with those of the authors invoking the critical pressure principle. It may be appreciated that in the patient Esophageal-4 classified with highly tense voicing by perceptual analysis presents the largest values for the BSC and the resulting esophageal voice is of very low quality showing a large noise component. On its turn the patient producing Esophageal-3 is on the opposite side showing very low BSC values resulting in aerial escape during the closure. The patient Esophageal-2 presenting the best results in the neo-glottal source and perceptual evaluations shows mean BSC values. These results avail the theory supporting the belief that for the production of esophageal voice independently of its quality a certain range of closure and muscular tenseness must be attained.

ACKNOWLEDGMENTS

This work is being funded by grants TIC2003-08756 and TEC2006-12887-C02-01/02 from Plan Nacional de I+D+i, Ministry of Education and Science, by grant CCG06-UPM/TIC-0028 from CAM/UPM, and by project HESPERIA (http.//www.proyecto-hesperia.org) from the Programme CENIT, Centro para el Desarrollo Tecnológico Industrial, Ministry of Industry, Spain.

REFERENCES

[1] Fernández-Baillo R., Fernández Camacho F.J., Nieto Altuzarra., Gómez Vilda P. Biomechanical Model of Voice Production. *Libro de Actas del XXIV Internacional Congress AELFA 2004*. Madrid. pp. 590.

[2] Isshiki N, Snidecor JC. Air intake and usage in esophageal speech. Acta Otolaryngol, vol 9, pp. 559-5574. 1965.

[3] Snidecor JC, Isshiki N. Air volume and air flow relationships of sic male asophageal speakers. J Speech Hear Disord, vol 30, pp. 205-216. 1965

[4] Gatenby RA., et all. Esophageal Speech: Double-contrast evaluation of the pharyngo-esophageal segment. Radiology, 157, pp. 127-131. 1985.

[5] Kilman WJ, Goyal RK. Disorders of pharyngeal and upper esophageal sphinter motor function. Arch Intern Med. Vol. 136, pp. 592-601.1976.

[6] Sloane PM., Griffin JM.., O'Dwyer TP. Esophageal insuffation and videofluroscopy for evaluation of esophageal speech in laryngectomy patients: clinical implications. Radiology, vol 181, pp. 433-437. 1991.

[7] Gómez, P., Fernández-Baillo., et al., Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters, Journal of Voice. Vol. 21, No. 4, 2007, pp. 450-476.
[8] Gonçalves MI ,Behlau M. Laringectomia total: perspectivas dareabilitaçao vocal. In Lopes Filho, O. Tratado de fonoaudiología.S„o Paulo, Roca, 1997.
[9] Vazquez de la Iglesia F, et all. Voz esofágica. Rev Med Univ Navarra. Vol 50, pp 56-64. 2006.
[10]Fernández-Baillo R., Gómez P. Métodos de análisis y evaluación acústica de la voz normal y patológica. In Proccedings The lenguaje of health care. Alicante. October. 2007.
[11] Fant G., Liljentcrants J., Lin Q., "A four-parameter model of glottal flow", STL-QSPR, Vol. 4, pp 1-13. 1985.
[12] Fernández-Baíllo, R., Gómez, P., Ramírez, C., Scola, B., “Pre-post surgery evaluation based on the profile of the glottal source”, Proc. of MAVEBA’07, Florence, December 13-15, 2007.
[13] Titze IR. Summary Statement. *Workshopon Acoustic Voice Analysis*. National Center of Voice and Speech. 1994.
[14] Frint VT, Pauka K. Erfahrungen mit der intra-osophagealen Druckmessung bei Laryngektomierten. Monatsschr Ohrenhell Laryngorhinol, 99:284-288. 1965.
[15] 11. Winans CS, Reichbach EJ, Waldrop WF. Esophageal determinants of alaryngeal speech. Arch Otolaryngol, 99:10-14. 1974.
[16] Aguiar-Ricz L et all. Behavior of the cricopharyngeal segment during esophageal phonation in laryngectomized patients. Journal of Voice, vol 21 (2), pp. 248-256. 2007.