EVIDENCE OF VOCAL CORD PATHOLOGY FROM THE MUCOSAL WAVE CEPSTRAL CONTENTS

P. Gómez, J. I. Godino, F. Rodríguez, F. Díaz, V. Nieto, A. Álvarez, V. Rodellar

DATSI, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28660 Boadilla del Monte, Madrid, Spain, e-mail:pedro@pino.datsi.fi.upm.es

ABSTRACT

The impact of voice pathologies caused by physiological alterations of the vocal cords is becoming a very important issue due to vocal abuse and unhealthy habits. Early detection of incipient damages to the cords may help in improving the prognosis, treatment and care of these pathologies. Information derived from the speech signal may help in detecting early stages of pathology, and to prevent them by assisting experts in voice therapy to correct vocal abuse in children and in professionals depending on voice as speakers, singers or lecturers among others. The present paper is devoted to detect the presence of certain pathologies in the voice or speech signal from the cepstral contents of the *mucosal wave* reconstructed by inverse filtering, using conclusions derived from the behavior of a 2-m vocal cord model. This can be of application in speech and sing education, and in pathology screening.

1. INTRODUCTION

The modelling of the vocal folds has been an old objective since the early works of Flanagan and Ishizaka [10] with a double purpose: to detect the vocal tract model and the glottal pulse pattern on one side, and to synthesize speech with a higher degree of naturalness on the other. Another important objective was to improve our understanding of the vocal folds operation during phonation both during normal and disordered voice production. A series of models devoted to explain the behavior of the vocal folds in 1-D, 2-D and 3-D have been produced during the last three decades which have completed our view of the phenomenon of voice production [13]. Under the purpose of the present work a model including the most relevant features have been used and adapted [8]: 2-mass assymetrical modelling, non-linear coupling between mass movement and glottal aperture, cord collision effects and non-linearities taken into account, defficient closure effects, lung flux excitation and vocal tract coupling effects. This is a modification of a well-known model through literature on the topic [3], [6]. The main characteristics of the model may be seen in Figure 1. The dynamic equations of this system are a set of four integro-differential equations, one for each of the masses in the system, similar to the one relative to mass M_{r1}:

$$f_{x1} - v_{r1}R_{r1} - M_{r1}\frac{dv_{r1}}{dt} - K_{r1}\int_{-\infty}^{t} v_{r1}dt - K_{r12}\int_{-\infty}^{t}(v_{r1} - v_{r2})dt = 0 \qquad (1)$$

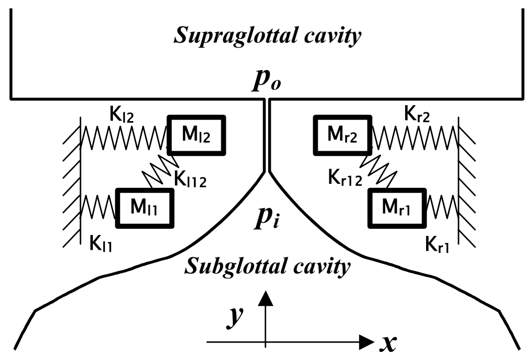

Figure 1. Schematic structure of the two-mass model.

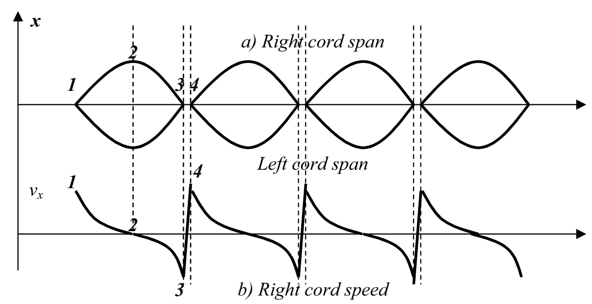

Figure 2. First vibration mode of the vocal cords. a) Glottal aperture. b) Right cord speed (unidimensional).

It is of most importance for our study to consider the behavior of the model with one mass per cord (with $M_{l2}=M_{r2}=0$ for example), as depicted in Figure 2. The vibration cycle starts at instant *1*, when both cords innitiate a fast separation (in the case of the right cord toward larger values of *x*). At instant *2* the right and left cords (considered symmetric) have arrived to their maximum span where the speed of the cords becomes zero. From this point on, the elastic forces restore the cords to their resting position, where at instant *3* both cords collide and bounce to *4*. The time that both cords remain in contact is very short, depending on the second order vibration of the model (known as *mucosal wave*). The intensity of the collision (the slope from *3* to *4*) is of special importance to infer if phonation is overstressed, pointing to damages in the vocal cords.

2. VOCAL CORD BEHAVIOR

The model can reproduce most of the features of normal or disordered voice, for instance Figure 3.a shows the glottal aperture synthesized under normal conditions, while its derivative may be seen in Figure 3.b. When the model is forced to operate with $M_{l2}=M_{r2}=0$, or when the mass-coupling stiffness is large enough the 2-mass model tends to behave as a 1-mass model, and the second vibration mode disappears, the vibration patterns (see Figure 4) showing good agreement with the conditions given in Figure 2. Other vibration modes are also possible depending on the parameter configurations, as commented for example in [12].

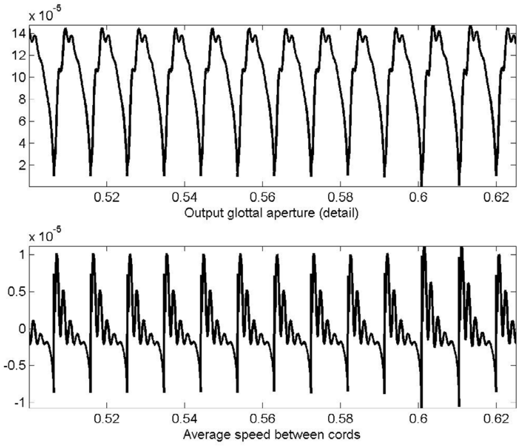

Figure 3. a) Glottal aperture for a symmetric distribution of parameters. b) Derivative of the glottal aperture.

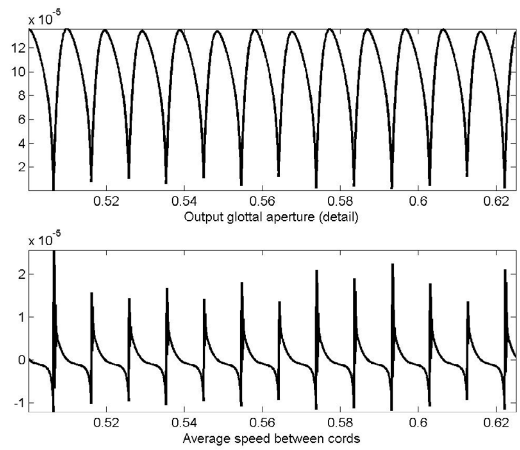

Figure 4. a) Glottal aperture for an excesive cord stiffness. b) Derivative of the glottal aperture.

3. GLOTTAL APERTURE RECONSTRUCTION

Through this section the reconstruction of the *mucosal wave* from the speech trace is sought following the work of Childers and others ([3], [2], [1]). The basics of inverse filtering as applied to this problem are explained in Figure 5.

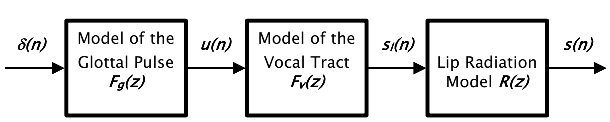

Figure 5. Voice production model.

The speech trace may be seen as the application of a train of delta pulses to a hypothetical glottal pulse generation model $F_g(z)$, its output u_g being modeled by the vocal tract transfer function $F_v(z)$ to yield voice at the lips s_l which is radiated as s. These operations may be seen in the time domain as the results of convolving the glottal and the vocal tract impulse responses f_g and f_v with trains of delta pulses, and transforming the results to radiated voice at the lips by the radiation model $r=\zeta^{-1}\{R(z)\}$:

$$s = \{\{\delta * f_g\} * f_v\} * r = \{f_g * f_v\} * r = s_l * r \qquad (2)$$

The inversion of this model starts by removing the radiation effects to get the radiation-compensated voice s_l:

$$s * h_r = \{s_l * r\} * h_r = s_l * \{r * h_r\} \cong s_l \qquad (3)$$

where it has been assumed that h_r is the inverse operator to r in the convolutional sense and obtaining a first estimation of the glottal inverse impulse response h_g, such that the influence of the glottal pulse may be removed from the radiation-compensated voice:

$$s_v = s_l * h_g \qquad (4)$$

The resulting *de-glottalized voice* s_v, may be seen as a first estimation of the impulse response of the vocal tract. It may be inverted finding its equivalent inverse impulse response h_{v0} (equivalent Wiener Filter), such that:

$$s_v * h_{v0} = \delta \qquad (5)$$

The inverse impulse response of the vocal tract may be used to remove the influence of the vocal tract from the radiation-compensated voice s_l by direct convolution producing a more accurate estimation of the glottal pulse u_g:

$$s_l * h_{v0} = u_{g0} \qquad (6)$$

These operations can be summarized as in Figure 6.

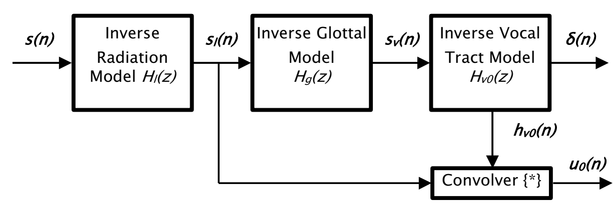

Figure 6. Estimation of the vocal tract inverse transfer function by coupled model inverter and convolver.

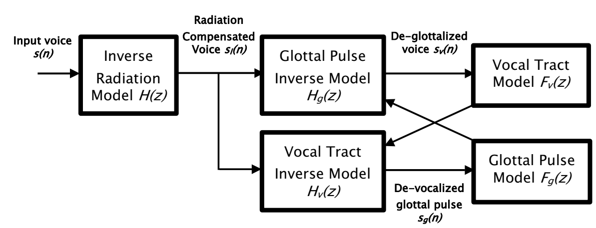

Figure 7. Implementation of the iteration scheme.

The new estimation of the glottal pulse may be used to model its inverse impulse response, which may be used to remove the glottal pulse influence from the radiation-compensated voice by convolution, producing a more accurate estimation of the de-glottalized voice, and so on. Through successive iterations as shown in Figure 7 a good estimation of both the glottal pulse and the de-glottalized voice may be obtained.

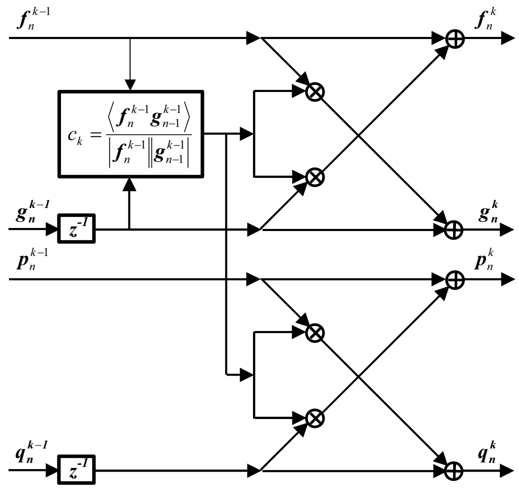

Figure 8. Coupled model inverter and convolver implemented by paired lattices.

Actually to carry out the described inversion processes the properties of lattice filters will be used. In fact, as shown for example in [4] the residual error from a lattice filter may be seen as the output of an all-pole filter inverse to the lattice input trace. This result allows to jointly build the inverse impulse response to $\boldsymbol{s_v}$ by a lattice reducing this signal to a white series, and at the same time convolve the associated signal $\boldsymbol{s_l}$ with the same inverse impulse response with a paired lattice using the same reflection coefficients as shown in Figure 8. The structure of the paired lattices is given in [9].

4. RESULTS

It is well known that the absence of *mucosal wave* is an indication of certain laryngeal pathologies [11]. The reconstruction of the glottal aperture may be used to detect the *mucosal wave*, as the second order vibration of the vocal cords. For such the above described algorithms have been applied to a normal segment of speech of 0.05 sec. duration corresponding to the vowel ***/a/*** as shown in Figure 9.a

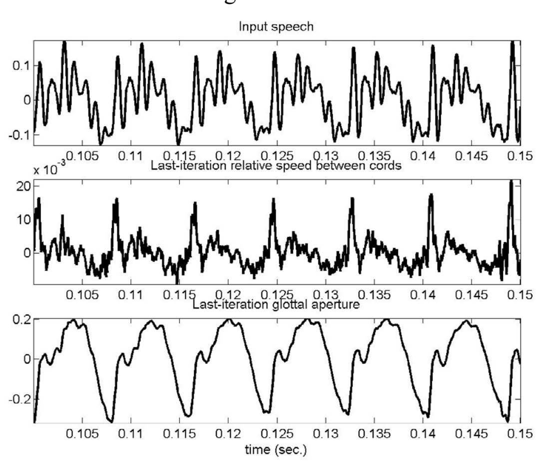

Figure 9. (Top to bottom). a) Input speech for a normal voice. b) Differential glottal aperture. c) Glottal aperture.

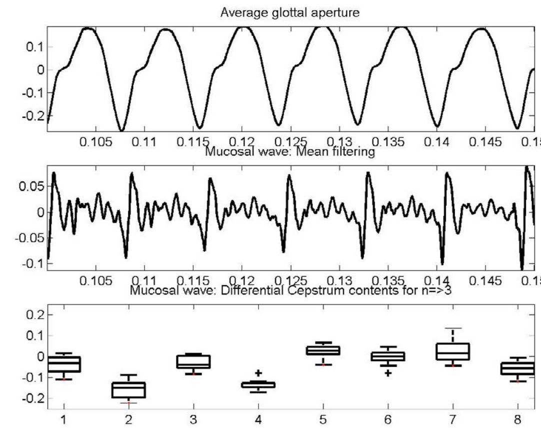

Figure 10. (Top to bottom). a) Mean component of the glottal aperture. b) Residual containing the *mucosal wave*. c) Cepstral contents after liftering and differentiation.

In Figure 9.b the differential glottal aperture resulting from applying the above described methods to the trace in Figure 9.a is shown. The integration of this last trace produces the glottal aperture in Figure 9.c, where the presence of a relatively large amount of *mucosal wave* may be observed as a ripple superposed to the main pattern of the aperture. The detection of the *mucosal wave* has been carried out in the present study applying first a median filter to the glottal aperture estimating the average value of the function on a *61* sample-wide sliding window:

$$\bar{u}_{gn} = \frac{1}{2K} \sum_{i=-K}^{K} u_{gn+i} \qquad (7)$$

subtracting this average from the trace, which produces a rough estimate of the *mucosal wave*:

$$w_{gn} = u_{gn} - \bar{u}_{gn} \qquad (8)$$

The average glottal aperture given in (7) is shown in Figure 10.a, whilst the estimation of the *mucosal wave* may be seen in Figure 10.b.

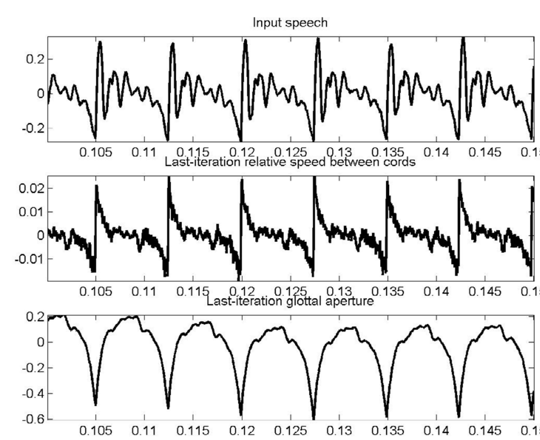

Figure 11. (Top to bottom). a) Input speech for a cord-stiffened pathological voice. b) Differential glottal aperture. c) Glottal aperture.

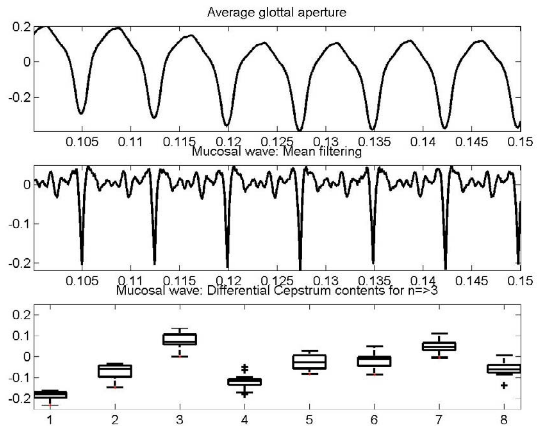

Figure 12. (Top to bottom). a) Mean component of the glottal aperture. b) Residual containing the *mucosal wave*. c) Cepstral contents after liftering and differentiation.

From the study of the envelope of the power spectrum of the *mucosal wave* it is concluded that the energy of the *mucosal wave* for normal voice concentrates mainly at lower frequencies and is larger than for pathological voice, therefore cepstral analysis may be used to detect the presence of pathology [4] using *512-samples* sliding windows. The cepstral coefficients $c_{n,m}$ are differenciated in m, the coefficients above the tenth being disregarded (liftering). The coefficient values and their statistical distributions are shown in Figure 10.c. These results have to be contrasted against others from speakers with pathologies affecting the presence of *mucosal wave*. A segment of 0.05 sec. of the vowel /a/ produced by a patient affected with vocal cord stiffness may be seen in Figure 11.a. The relative speed between cords is shown in Figure 11.b, adhering to the theoretical previsions and to the results of the model forced to vibrate following the first mode (see Figure 2 and Figure 4). The reconstruction of the glottal aperture given in Figure 11.c shows that the presence of mucosal vibration is much less strong than in the normal case. The application of the median filter results in the trace given in Figure 12.a, and the *mucosal wave* estimation is given in Figure 12.b. The plot in Figure 12.c shows the first eight cepstral coefficients and their statistical distributions.

5. CONCLUSIONS

From the results shown it seems evident that the sequence of the first differential cepstral coefficients reflects the presence of *mucosal wave*. In general the presence of a strong *mucosal wave* is correlated with a faster alternating pattern in the lower differential cepstral coefficients, while a slower fluctuation is associated with a lower amount of *mucosal wave*. The presence of *mucosal wave* can be clearly established visually on the glottal aperture. The quantization of this presence and its frequency distribution may be of great help in the detection of vocal fold pathologies. A more detailed study should have to be conducted based in Principal Component Analysis to establish which cepstral coefficients are more relevant to the study for descriptors of pathologies, which could be used in data analysis and evaluation to support diagnose aids by automatic pattern classifiers, as for example Neural Nets or Gaussian Mixture Models.

6. ACKNOWLEDGMENTS

This research is being carried out with the support of the following grants from the Programa Nacional de las Tecnologías de la Información y las Comunicaciones: TIC-2002-0273, TIC2003-08956-C02-00, TIC2003-08756.

7. REFERENCES

[1] Alku, P., Vilkman, E., "Estimation of the glottal pulseform based on Discrete All Pole modeling", *Proceedings of the ICSLP*, 1994, pp. 1619-1622.

[2] Cheng, Y. M., O'Shaughnessy, D., "Automatic and reliable estimation of the glottal closure instant and period", *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 37, 1989, pp. 1805-1815.

[3] Childers, D. G., *Speech Processing and Synthesis Toolboxes*, John Wiley & Sons, New York, 2000.

[4] Deller, J. R., Proakis, J. G., and Hansen, J. H. L., *Discrete-Time Processing of Speech Signals*, Macmillan Pub. Co., Englewood Cliffs, NJ, 1993.

[5] Deng, H., Beddoes, M., Ward, R., Hodgson, M., "Estimating the Vocal-Tract Area Function and the Derivative of the Glottal Wave from a Speech Signal", *Proc. of the EUROSPEECH'03*, 2003, pp. 2437-2440.

[6] Giovanni, A., Ouaknine, M., Guelfucci, B., Yu, P., Zanaret, M. and Triglia, J. M., "Nonlinear Behavior of Vocal Fold Vibration: The Role of Coupling Between the Vocal Folds", *Journal of Voice*, Vol. 13, No. 4, 1999, pp. 465-476.

[7] Gómez, P., "Fundamentals of Glottal Pulse Extraction by Inverse Filtering of the Speech Trace", Research Report, project MAPACI ref. TIC-2002-02273, Facultad de Informática, Universidad Politécnica de Madrid, Madrid, Spain, March 2003.

[8] Gómez, P., "Fundamentals of the Electromechanical Modelling of the Vocal Tract", Research Report, project MAPACI ref. TIC-2002-02273, Facultad de Informática, Universidad Politécnica de Madrid, Madrid, Spain, March 2003.

[9] Gómez, P., Nieto, V., Martínez, R., Álvarez, A., Rodríguez, F., Rodellar, V., "Detection of the *Mucosal wave* in the Speech Signal for Larynx Pathology Characterization", *Proceedings of the International Symposium on Information and Communication Technologies*, Ed. Trinity College Dublin, Ireland, September 2003, pp. 33-39.

[10] Ishizaka, K., Flanagan, J. L., "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell Systems Technical Journal*, Vol. 51, 1972, pp. 1233-1268.

[11] Rydell, R., Schalen, L., Fex, S., Elner, A., "Voice evaluation before and after laser excision vs. Radiotherapy of t1A glottic carcinoma", *Acta Otolaryngol.*, Vol. 115, No. 4, 1995, pp. 560-565.

[12] Sciamarella, D., d'Alessandro, C., "Reproducing laryngeal mechanisms with a two-mass model", *Proc. of the EUROSPEECH'03*, 2003, pp. 2385-2388.

[13] Titze, I. R., *Principles of Voice Production*, Prentice-Hall, Englewood Cliffs, N. J., 1994.