A new mucosal wave correlate detection method as a clue to voice pathology

Gómez, P., Godino, J. I., Díaz, F., Martínez, R., Nieto, V., Rodellar, V.

Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28660 Madrid Spain Tel: +34.913.367.384, Fax: +34.913.366.601
e-mail: pedro@pino.datsi.fi.upm.es

Abstract

Voice recordings may be the only sources for pathology detection in cases when the use of invasive instrumentation is not feasible or recommended as in new-borns and long-distance screening, among others. Classical pathology detection methods using voice rely on processing basic information from the voice signal, as the *pitch*, *jitter*, *shimmer*, *HNR* and others similar. In the present work a new method to estimate a measure related to *HNR* from the detection and processing of a signal correlate of the *mucosal wave* is presented as the classical definition of *HNR* is rather vague. An evaluation of *mucosal wave* in recordings from normal and pathological cases is presented and discussed. The method is validated contrasting evidence derived from glottal dynamics.

1. Introduction

Through the present work the precise reconstruction of a *mucosal wave correlate* from voice using inverse filtering of real and simulated traces is presented. This signal is of most importance in establishing the presence of certain pathologies in the vocal folds [8]. In what follows the term *mucosal wave* will refer to the travelling wave effect taking place in the vocal cords due to the distribution of masses on the *body cover* and related tissues, and the term *mucosal wave correlate* (*MWC*) will be used for the influence of *mucosal wave* on the overall pattern of the *glottal source derivative (GSD)*, appearing as a superimposed waveform on this trace. The *MWC* may be seen as a higher order vibration regime of the vocal folds, once the average main movement or first regime has been removed. For this study a version of the *vocal cord 2-mass model* as given in [7] has been implemented in MATLAB® [5], its main features being: 2-mass asymmetric modelling, non-linear coupling between mass movement and *GSD*, cord collision effects, non-linearities and deffective closure effects, lung flux excitation and vocal tract coupling. The parameters of the model are the *lumped masses (2 per cord)* M_{1l} and M_{2l} (left cord), M_{1r} and M_{2r} (right cord), the *elastic parameters* K_{1l} and K_{2l} (relative to reference) and K_{12l} (intercoupling), and their respective ones for the right cord: K_{1r}, K_{2r} and K_{12r}. The dynamic equations of the model are a set of four integro-differential equations, one for each of the masses in the system, with the following structure:

$$f_{xi,j} - v_{i,j}R_{i,j} - M_{i,j}\frac{dv_{i,j}}{dt} - K_{i,j}\int_{-\infty}^{t} v_{i,j}\,dt - \\ - K_{12i,j}\int_{-\infty}^{t}(v_{1j} - v_{2j})\,dt = 0 \qquad (1)$$

where $i \in \{1,2\}$ determines the subglottal (*1*) or the supraglottal (*2*) cords and $j \in \{l,r\}$ distinguishes left from right cords, $f_{xi,j}$ is the force acting on the cord in the direction of the axis x (transversal) resulting from the action of the pressure difference between the subglottal and supraglottal regions p_i-p_0 (the excitation), and $v_{i,j}$ is the corresponding mass speed along the axis x (the response). Although this is a simplification of more elaborated models [9], it reproduces the vibration features of interest for the present study.

2. Modelling cord movement

Adopting standard values [7] for the parameters in the model an example of the resulting *GSD* may be seen in Figure 1 (Top). Due to the difference between the values of $M_{1r,l}$ and $M_{2r,l}$ the subglottal masses will describe a pattern of movement approaching a rectified sinusoid. On its turn the supraglottal masses will describe a more complicated pattern of movement due to interactions against the reference and the massive subglottal masses [2].

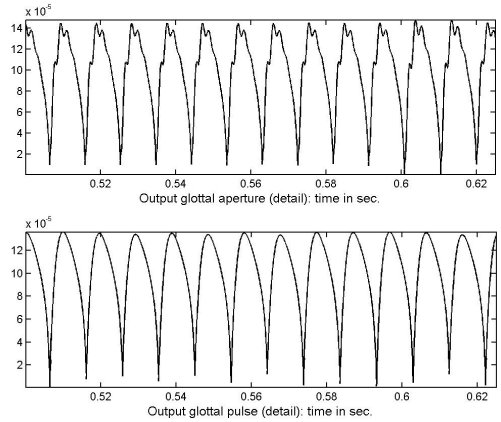

Figure 1. Top: Simulated *GSD* for normal voice conditions. Bottom: Idem for modeled cord-stiff pathology.

The *GSD* may be seen as the aggregation of two vibration components: *the slow and long range component (SLRC)* due mainly to subglottal masses, and the *fast and small range component (FSRC)*, which reflects *supraglottal mass* movement appearing as the over-ringing on the signal. This is most plausibly associated with the *mucosal wave*, as its presence is due to the coupling between *lower and upper masses*, otherwise the whole cord would act as a single mass, as revealed in when the supraglottal mass is removed or the intercoupling $K_{12i,j}$ is extremely high (stiff cord) as shown in Figure 1 (Bottom), where no traces of the *MWC* are apparently found.

3. Estimating the *GSD*

The reconstruction of the *MWC* from the voice trace is based in inverting ([3], [1]) the well-known voice production model given in for instance in [4], pp. 193. The voice trace $\boldsymbol{s}$ may be seen as the output of a generation model $\boldsymbol{F_g(z)}$ excited by a train of delta pulses, its output being modelled by the vocal tract transfer function $\boldsymbol{F_v(z)}$ to yield voice at the lips $\boldsymbol{s_l}$ which

is radiated as $\boldsymbol{s}$, where $\boldsymbol{r=\zeta^{-1}\{R(z)\}}$ is the radiation model and $\boldsymbol{f_g}$ and $\boldsymbol{f_v}$ are the glottal and vocal tract impulse responses:

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

This model will be inverted to reconstruct the *GSD* $\boldsymbol{u=\delta * f_g}$ from the voice trace $\boldsymbol{s}$ by removing the radiation effects to get the radiation-compensated voice $\boldsymbol{s_l}$. A first estimation of the *Inverse Glottal Impulse Response* $\boldsymbol{h_g}$ may be used to reconstruct the *de-glottalized voice* $\boldsymbol{s_v}$, from which a first estimation of the *Inverse Vocal Tract Impulse Response* $\boldsymbol{h_{v0}}$ may be derived, which may be used to remove the influence of the vocal tract from the radiation-compensated voice $\boldsymbol{s_l}$ by direct convolution producing a first estimation of the *glottal source* $\boldsymbol{u_0}$ as summarized in Figure 2.

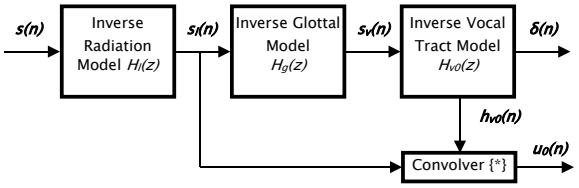

Figure 2. Estimation of the *glottal source* $\boldsymbol{u_0}$ by a coupled model inverter and convolver.

Through recursion good estimates of both the *glottal source* at iteration step *i*, $\boldsymbol{u_i}$ and its derivative $\boldsymbol{u_{gi}}$ (*GSD*) may be obtained. The algorithmic details of this procedure are given in [6].

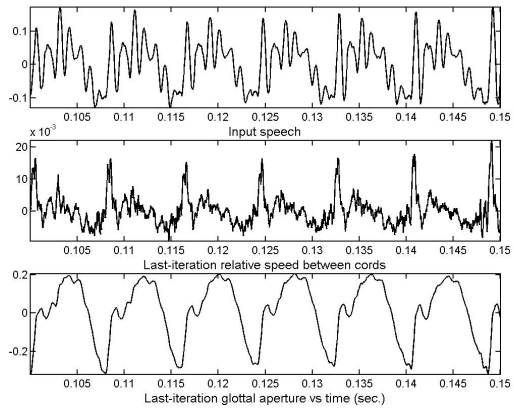

Figure 3. Normal voice. Top: Input voice. Middle: Differential *GSD*. Bottom: *GSD*.

The described procedure has been applied to a trace of non-pathological voice corresponding to the vowel **/a/,** of which a segment of 0.05 sec. of duration is shown in Figure 3.

4. Estimating the *MWC*

Several techniques were used to remove the *SLRC* and produce an estimate of the *MWC*, as mean-, low pass- and cepstral filtering [6], showing good results when the *GSD* minima are not too sharp, otherwise the residual component of the *SLRC* near the minima is large and it distorts the *MWC* estimate. The technique proposed to solve this problem is based on a period-by-period subtraction of the slow-moving baseline on which the minima of the *GSD* relies, and on DFT low-pass filtering. The first process is explained in Figure 4. The method is based in subtracting the slope joining two successive minima in the *GSD* ($u_{gmin,k-1}$ and $u_{gmin,k}$) and the corresponding increment $\Delta u_{g,n}$ at time instant *n* from the *GSD*:

$$u_{gf,n} = u_{g,n} - \Delta u_{g,n} =$$

$$= u_{g,n} - \frac{u_{g\,min,k} - u_{g\,min,k-1}}{m_k - m_{k-1}}\left(n - m_{k-1}\right) \quad (3)$$

$$u_{gm,n} = u_{gf,n} - \min\left\{u_{gf,n}\right\} \quad (4)$$

$$u_{gu,n} = (-1)^k u_{gm}\left(n \in w_k\right) \quad (5)$$

where w_k is the *k-th cycle window*. The signal $u_{gu,n}$ could be seen as half the excursion that one of the vocal cords would describe if vibrating freely (no opposite cord).

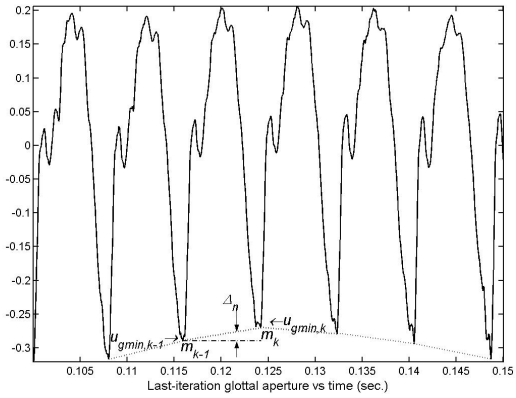

Figure 4. Levelling method used for *GSD*.

By subtracting the minima of *(3)* and sign reversal of each alternate cycle-window w_k *(5)* the *unfolded GSD* is obtained (see Figure 5).

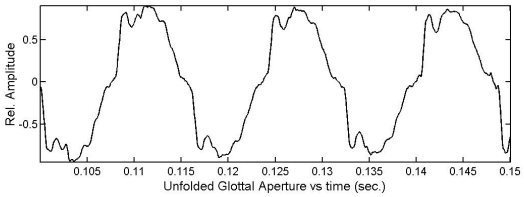

Figure 5. Unfolded *GSD* $u_{gu,n}$.

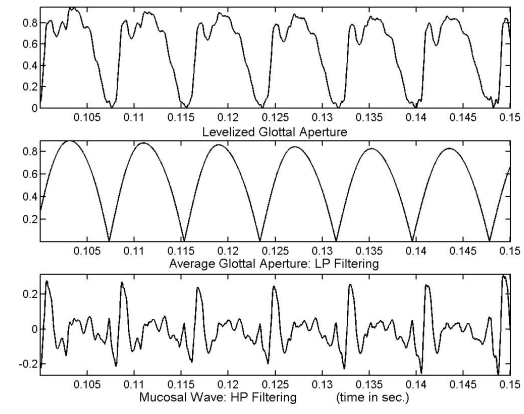

Figure 6. Normal voice. Top: Levelled *GSD*: $u_{gf,n}$. Mid: Low-pass filtered *GSD*: $u_{gl,n}$. Bottom: High-pass filtered *GSD*: $u_{gh,n}$.

The *unfolded GSD* can be low-pass filtered using spectral truncation in the frequency domain by means of the DFT:

$$U_{gb}(m) = W_{lp}(m) \sum_{n \in w_k} u_{gu,n} e^{-jm\frac{2\pi}{N_k}n} \quad (6)$$

$$u_{gl,n} = \left|u_{gb,n}\right| = \left|\frac{1}{N_k}\sum_{m=0}^{N_k-1} U_{gb}(m)e^{jn\frac{2\pi}{N_k}m}\right| \quad (7)$$

$$u_{gh,n} = u_{gb,n} - u_{gl,n} \quad (8)$$

where $W_{lp}(m)$ is a low-pass window and N_k is the size of the *k-th cycle* window. The low-frequency trace $u_{gl,n}$ is obtained by inverse DFT and rectification (7), and the high-frequency trace $u_{gh,n}$ by subtraction (8), these traces shown in Figure 6.

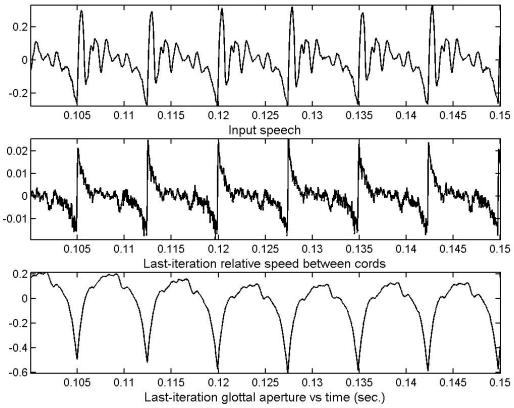

Figure 7. Pathological voice (stiffness): Top: Input voice. Middle: Differential *GSD*. Bottom: *GSD*.

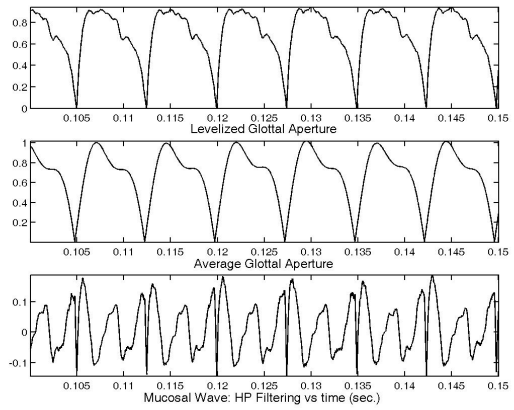

Figure 8. Pathological voice: Top: *Levelled GSD*. Mid: Low-pass filtered *GSD*. Bottom: High-pass filtered *GSD*.

The results of estimating the *GSD* for cord-stiff pathological voice (*/a/*) are given in Figure 7, and the results after levelling, low-pass filtering and subtraction are shown in Figure 8 (Bottom). It may be seen that the size of the *MWC* is smaller in this case than in normal voice (Figure 6, Bottom).

5. Results and Discussion

In Figure 9 a comparison between the results from normal vs pathological voice is given, using the power ratio (related to *HNR*) between the *FSRC* and *SLRC*:

$$r_{pk} = \frac{\sum_{n\in w_k}\left[u_{gb}(n) - u_{gl}(n)\right]^2}{\sum_{n\in w_k} u_{gl}^2(n)} \quad (9)$$

where w_k is the *k-th* period-adjusted window used in the evaluation of r_{pk}, u_{gb} is the *levelled GSD*, and u_{gl} is the *SLRC*.

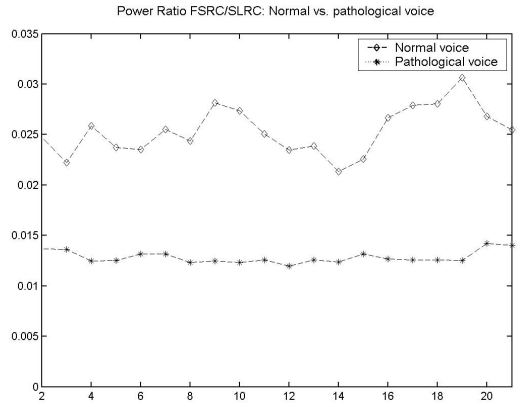

Figure 9. Power ratio between *FSRC* and *SLRC*: normal vs. stiff-pathological voice for consecutive cycle-window frames.

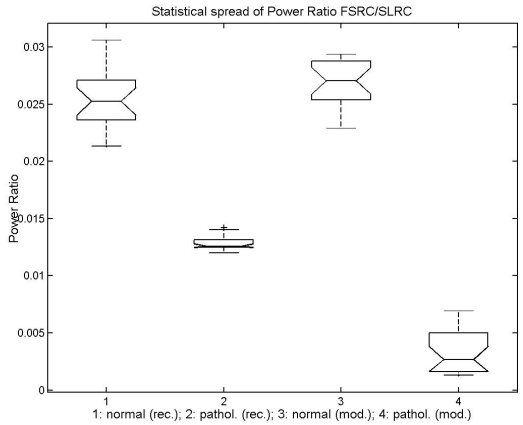

Figure 10. Statistical dispersion of the power ratio between *FSRC* and *SRLC* for normal and pathological voice from recordings (rec.) and simulations (mod.).

The differences in r_p between normal and stiff-cord voice are due to the amount of *MWC* present in both cases. To check this, a set of traces was produced using the vocal cord model with the following settings for normal voice: $M_{l,r1}$*=0.2 g*, $M_{l,r2}$*=0.02 g*, $K_{l,r1}$*=40,000 dyn/cm*, $K_{l,r2}$*=100,000 dyn/cm*, $K_{l,r12}$*=30,000 dyn/cm*, and the same parameter values for pathological voice with the exception: $M_{l,r2}$*=0 g*. The plot in Figure 10 compares the values of r_p among traces used in the study (normal vs pathological, recorded vs simulated), showing the dispersion of this parameter for the estimation windows used. It may be concluded that the ratio r_p for normal voice is larger than for pathological voice. The results for normal voice from recordings and from simulations compare within the same ranges. As simulation results may be adjusted using the model settings, the biomechanical parameters may be obtained through model parameter adaptation for a real case under exploration. Besides, the confidence intervals for pathological and non-pathological traces do not overlap, showing considerable separation gaps, concluding that this parameter (r_p between *SLRC* and *FSRC*) could be a good distortion measurement for cord stiffness pathology detection. It may be interesting to consider now if the proposed extraction method is realistic, yielding results which are in agreement with the physical reality of vocal cord dynamics, assessing if what is called *mucosal wave correlate* is really related to the physical essence of the *mucosal wave*.

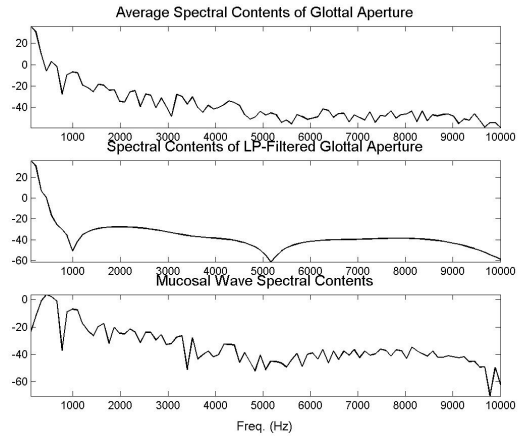

Figure 11. Typical power spectral densities of *Levelled GSD* (top), Low-pass filtered *GSD* (midd.) and High-pass filtered *GSD* (*mucosal wave correlate*) for normal voice.

For such, it is of interest to compare the spectral power density of *the mucosal wave correlate* obtained from a specific case of normal voice as given in Figure 11 (bottom), against the spectral pattern of the modulus of the *vocal cord transadmittance* obtained from the *2-mass* model in the frequency domain, associating the force f_1 and f_2 acting on both masses of the same cord (as for instance M_{1r} and M_{2r}) with the observable mass velocities v_1 and v_2 in the electromechanical equivalent to the one-cord dynamical system given in Figure 12, derived from the respective dynamical equations given in (1) where forces play the role of electromotive forces and velocities that of currents. Details of this study are given in full length in [10].

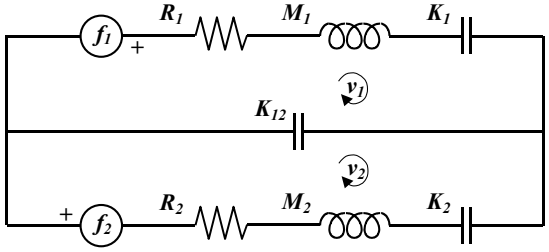

Figure 12. Equivalent electromechanical circuit of the right cord in the two-mass model.

When plotting the *cord trans-admittance* relating f_1 and v_1 in the frequency domain, using *ad hoc* values for the biomechanical parameters of the model, a curve as the one given in Figure 13 may be found to optimally match the real case given in Figure 11, showing a notch between two maxima for frequencies below 1000 Hz which greatly resemble the general pattern of the *MWC* power spectral density.

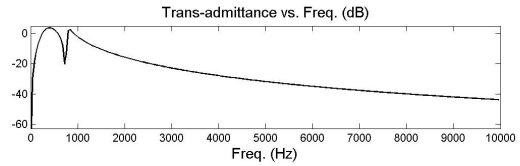

Figure 13. Square modulus of the *vocal cord trans-admittance* for a set of biomechanical parameters matching the *mucosal wave correlate* power spectral density in Figure 11 (bottom).

This behavior or a similar one may be found in most of normal voice cases studied where a *2-mass* mode of vibration is established during phonation. It may be shown that maxima sizes and positions are influenced by the value of both masses, and the notch position and size is mainly influenced by the intercoupling between masses (K_{12}). Curve fitting may grant optimal detection of biomechanical parameters of the system best approaching each case under study. For systems which show several notches encapsulated between maxima, a model may be established incorporating a pair of coupled masses for each notch. The sharper the notch, the looser the coupling. Stiffer *2-mass* systems will show shallower notches or no notch at all. A new fingerprint for the *mucosal wave* may be established using the number of notches and surrounding peaks, the position and value of the peaks, and the position and value of the notch, which may be used in speech pathology detection and classification, and for speaker identification as well, possibly opening new lines of research.

6. Acknowledgments

This research is carried out under the following projects of Programa Nacional de las Tecnologías de la Información y las Comunicaciones from the Ministry of Science and Technology of Spain: TIC-2002-0273, TIC2003-08956-C02-00, TIC2003-08756.

7. References

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

[2] Berry, D. A., "Mechanisms of modal and nonmodal phonation", *Journal of Phonetics*, Vol. 29, 2001, pp. 431-450.

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

[4] Deller, J. R., Hansen, J. H. L., Proakis, J. G., *Discrete-Time Processing of Speech Signals*, John Wiley & Sons, New York, 2000.

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

[6] Gómez, P., Godino, J. I., Rodríguez, F., Díaz, F., Nieto, V., Álvarez, A., Rodellar, V., "Evidence of Vocal Cord Pathology from the Mucosal Wave Cepstral Contents", *Proc. of the ICASSP'04*, Montreal, Canada, May 17-21, 2004, pp. V.437-440.

[7] 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.

[8] 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.

[9] Story, B. H., and Titze, I. R., "Voice simulation with a bodycover model of the vocal folds", *J. Acoust. Soc. Am.*, Vol. 97, 1995, pp. 1249–1260.

[10] Gómez, P., Godino, J. I., Díaz, F., Álvarez, A., Martínez, R., Rodellar, V., "Biomechanical Parameter Fingerprint in the Mucosal Wave Power Spectral Density", *Proc. of the ICSLP'04,* Jeju Island, South Korea, October 4-8, 2004 (to appear).