# PRECISE RECONSTRUCTION OF THE MUCOSAL WAVE FOR VOICE PATHOLOGY DETECTION AND CHARACTERIZATION

*P. Gómez, F. Díaz, R. Martínez, J. I. Godino, A. Álvarez, F. Rodríguez, V. Rodellar*

Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28660 Boadilla del Monte, Madrid, Spain
phone: +34.91.3367384, fax: +34.91.3366601, e-mail: pedro@pino.datsi.fi.upm.es

## ABSTRACT

Voice recordings are the only basis for pathology detection and classification in critical cases where invasive instrumentation is not possible 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 *HNR* from the detection and processing of a signal correlate of the *mucosal wave* is presented, as it is well known that *mucosal wave* alterations give clues to the presence of certain pathologies. An evaluation of *mucosal wave* in recordings from normal and pathological cases is presented and discussed, checking the results against those produced from simulated voice by a *2-mass model*.

## 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 aperture*, appearing as a superimposed ringing on its reconstructed 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. To start the study a version of the *vocal cord 2-mass model* as given in [7] and [4] has been implemented in MATLAB® [5], its main features being: 2-mass asymmetric modelling, non-linear coupling between mass movement and glottal aperture, cord collision effects, non-linearities and deffective closure effects taken into account, 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:

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

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). The effective *glottal aperture* resulting from the position of the subglottal and supraglottal ridges (known also as *lower and upper lips*), will establish the flux of air induced in the vocal tract, and has been defined as:

$$u_g = (p_i - p_o)\frac{L}{\rho c}\frac{\int_{-\infty}^{t}(v_{1r} - v_{1l})dt\int_{-\infty}^{t}(v_{2r} - v_{2l})dt}{\int_{-\infty}^{t}(v_{1r} + v_{2r} - v_{1l} - v_{2l})dt} \qquad (2)$$

where *L* is the equivalent length of the vocal cords, $\rho$ is the density of air and *c* is the speed of sound. The movement of the vocal cords, will be described by the *glottal aperture* and its derivative (*cord relative speed*).

## 2. MODELLING CORD MOVEMENT

Adopting standard values [7] for the parameters in the model the resulting *glottal aperture* may be seen in Figure 1. Due to the difference between the values of $M_{1r,l}$ and $M_{2r,l}$ the subglottal masses will move more inertially, describing a pattern 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].

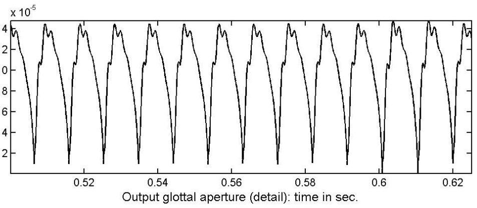

**Figure 1.** Simulated *glottal aperture* under normal conditions.

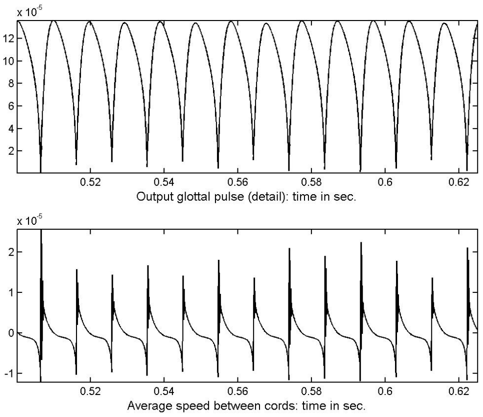

**Figure 2.** Top: *Glottal aperture* for modeled cord stiffness. Middle: Derivative of the *glottal aperture*.

The *glottal aperture* may be seen as the aggregation of the two vibration orders: *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 in Figure 1. 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. Although this is a simplification of more elaborated models [9], it reproduces the vibration features of interest for the present study. To illustrate this the *glottal aperture* produced by the *2-mass* model when the sub-supraglottal stiffness parameters $K_{12l,r}$ have been substantially increased is given in Figure 2. It may be seen that the *FSRC* apparently vanishes out. An ideal study of this behavior from a simple *1-mass* system is given in Figure 3.

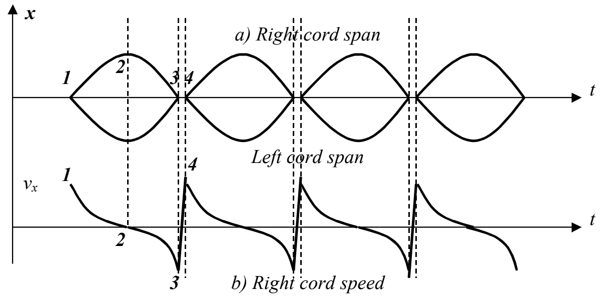

**Figure 3.** First vibration mode of the vocal cords. Top: Glottal aperture. Bottom: Right cord speed (unidimensional).

The vibration cycle starts at instant *1*, when both cords initiate a fast separation. At instant *2* the right and left cords 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 separate again with opposite velocities (instant *4*). The intensity of the collision (the slope from *3* to *4*) is of special importance to measure overstress in phonation.

## 3. ESTIMATING THE GLOTTAL APERTURE

The reconstruction of the *MWC* from the voice trace is based in inverting ([4], [3], [1]) the well-known voice production model given in for instance in [4], pp. 193. The voice trace ***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 ***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:

$$s=\{\{\delta * f_g\} * f_v\} * r=\{f_g * f_v\} * r=s_l * r \qquad (3)$$

This model will be inverted to reconstruct the glottal aperture $\boldsymbol{u=\delta * f_g}$ from the voice trace ***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 pulse* $\boldsymbol{u_0}$ as summarized in Figure 4.

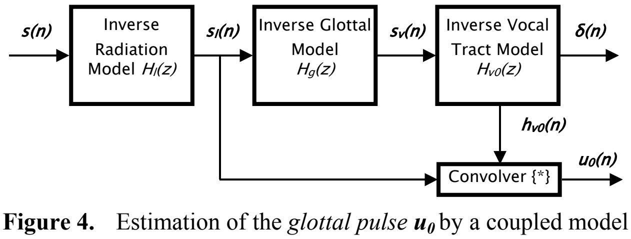
s(n)
Inverse Radiation Model $H_l(z)$
$s_l(n)$
Inverse Glottal Model $H_g(z)$
$s_v(n)$
Inverse Vocal Tract Model $H_{v0}(z)$
δ(n)
$h_{v0}(n)$
$u_0(n)$
Convolver {*}

**Figure 4.** Estimation of the *glottal pulse* $\boldsymbol{u_0}$ by a coupled model inverter and convolver.

From $\boldsymbol{u_0}$ a new *Inverse Glottal Impulse Response* $\boldsymbol{h_{g1}}$ may be produced to remove the *glottal pulse* influence from the radiation-compensated voice $\boldsymbol{s_{l1}}$, producing a more accurate estimation of the de-glottalized voice $\boldsymbol{s_{v1}}$, and this on its turn may be inverted to remove vocal tract influence from the radiation compensated voice and re-estimate the *glottal pulse* $\boldsymbol{u_1}$, and so on. Through recursion good estimates of both the *glottal pulse* at iteration step *i*, $\boldsymbol{u_i}$ and its integral $\boldsymbol{u_{gi}}$ (*glottal aperture*) may be obtained. The basics and algorithmic details of this procedure are given in [6].

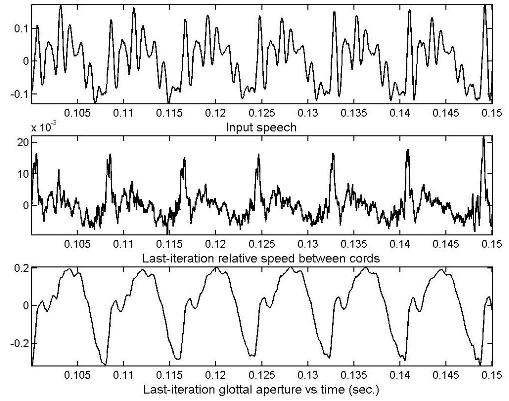

**Figure 5.** Normal voice. Top: Input voice. Middle: Differential *glottal aperture*. Bottom: *Glottal aperture*.

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 5.

## 4. ESTIMATING THE *MWC*

To remove the *SLRC* and produce an estimate of the *MWC* several techniques were used, as mean-, low pass- and cepstral filtering [6], showing good results when the *glottal aperture* minima (glitches) are not too sharp, otherwise the residual component of the *SLRC* near the minima is large and it distorts the pattern of 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 glottal aperture relies, and on DFT low-pass filtering. The first process is explained in Figure 6.

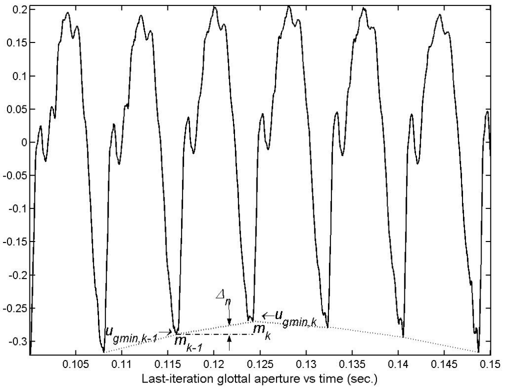

**Figure 6.** Levelling method used for *glottal aperture*.

The method is based in estimating the slope joining two successive minima in the *glottal aperture* ($u_{gmin,k-1}$ and $u_{gmin,k}$) and the corresponding increment $\Delta u_{g,n}$ at time instant *n* which should be subtracted from the glottal aperture:

$$u_{gf,n}=u_{g,n}-\Delta u_{g,n}=u_{g,n}-\frac{u_{gmin,k}-u_{gmin,k-1}}{m_k-m_{k-1}}\left(n-m_{k-1}\right) \qquad (4)$$

$$u_{gm,n}=u_{gf,n}-min\left\{u_{gf,n}\right\} \qquad (5)$$

$$u_{gu,n}=\left(-1\right)^k u_{gm}\left(n\in w_k\right) \qquad (6)$$

where $w_k$ is the *k-th period 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). By subtracting the minima of (4) and reversing the sign of each alternate period-window $w_k$ (6) the *unfolded glottal aperture* will be obtained (see Figure 7).

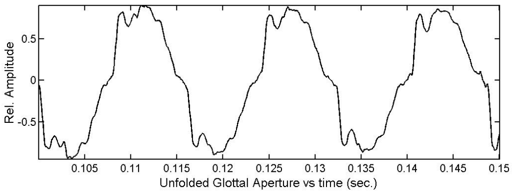

**Figure 7.** Unfolded *glottal aperture* $u_{gu,n}$.

This signal can be now low-pass filtered using spectral truncation in the frequency domain by means of the DFT:

$$U_{gb}(m)=W_{lp}(m)\sum_{n\in w_k}u_{gu,n}e^{-jm\frac{2\pi}{N_k}n} \qquad (7)$$

$$u_{gl,n}=\left|u_{gb,n}\right|=\left|\frac{1}{N_k}\sum_{m=0}^{N_k-1}U_{gb}(m)e^{jn\frac{2\pi}{N_k}m}\right| \qquad (8)$$

$$u_{gh,n}=u_{gb,n}-u_{gl,n} \qquad (9)$$

where $W_{lp}(m)$ is a low-pass window in the frequency domain and $N_k$ is the size of the *k-th period window*. The low-frequency trace $u_{gl,n}$ is obtained by inverse DFT and rectification (8), and the high-frequency trace $u_{gh,n}$ by subtraction (9), the results being shown in Figure 8. Their corresponding power spectra are given in Figure 9.

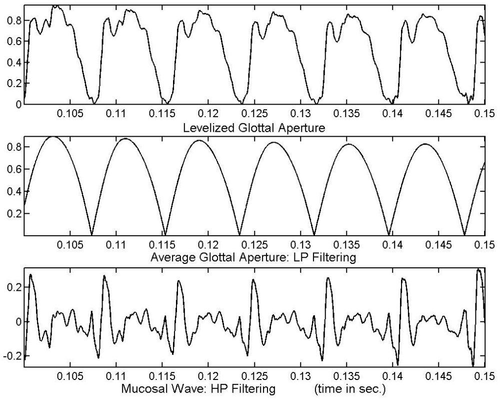

**Figure 8.** Results for normal voice from recordings. Top: Levelled signal $u_{gf,n}$. Middle: Low-pass filtered *glottal aperture* $u_{gl,n}$ (*SLRC*). Bottom: High-pass filtered *glottal aperture* $u_{gh,n}$ (*FSRC*).

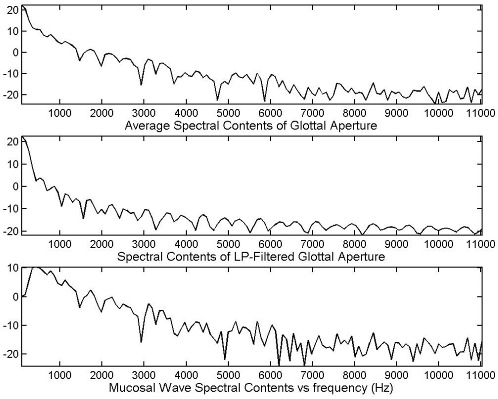

**Figure 9.** Spectral contents of traces in Figure 8.

The results of estimating the *glottal aperture* for cord-stiff pathological voice (*/a/*) are given in Figure 10, and the corresponding ones for levelling, low-pass filtering and subtraction are shown in Figure 11. It may be seen from the bottom template that the amount of *FSRC* seems to be smaller for this case than for normal voice (Figure 8, bottom).

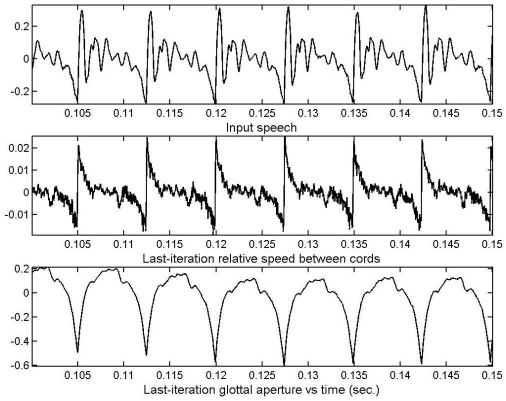

**Figure 10.** Pathological voice from recordings: Top: Input voice. Middle: Differential *glottal aperture*. Bottom: *Glottal aperture*.

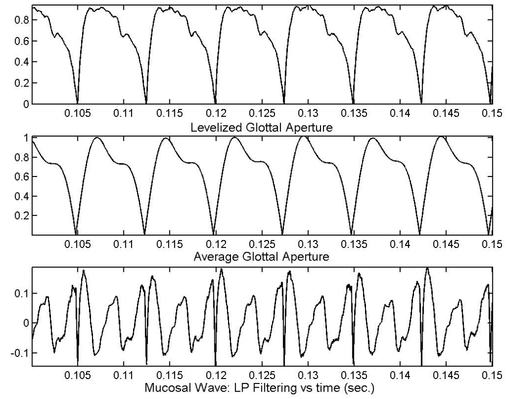

**Figure 11.** Results for pathological voice from recordings: Top: *Levelled glottal aperture*. Middle: Low-pass filtered correlate of the *SLRC*. Bottom: High-pass filtered correlate of the *FSRC*.

## 5. RESULTS AND DISCUSSION

In Figure 12 a comparison between the results from normal vs pathological voice is given, using the power ratio *(HNR)* between the *FSRC* and *SLRC*:

$$r_{pk}=\frac{\sum_{i\in w_k}\left[u_{gb}(n)-u_{gl}(n)\right]^2}{\sum_{i\in w_k}u_{gl}^2(n)} \qquad (10)$$

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

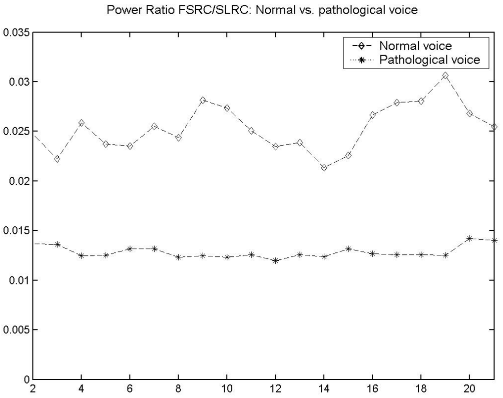

**Figure 12.** *HNR* Ratio: normal vs. stiff-pathol. voice (recordings).

The difference in the *HNR* measured from normal and stiff-cord voice is due to the different amount of *MWC* present in both cases. To check this hypotesis a new set of traces was produced from 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*. Normal voice traces from simulations are given in Figure 13.

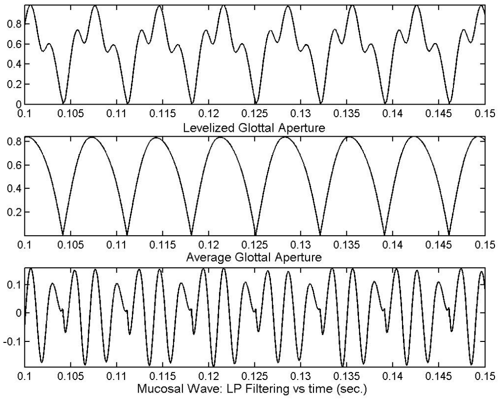

**Figure 13.** Traces from simulations (normal case): Top: Levelled signal. Middle: Low-pass filtered correlate: *SLRC*. Bottom: High-pass filtered correlate: *FSRC*.

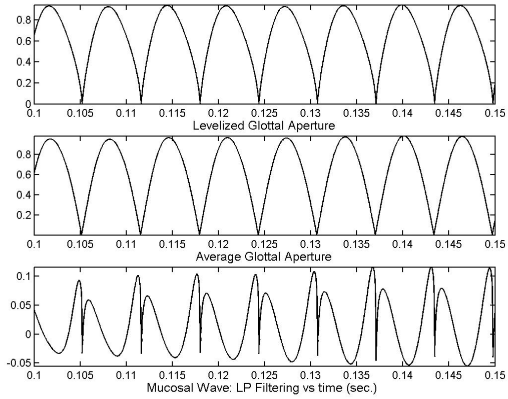

**Figure 14.** Traces from simulations (stiff-pathological case): Top: Levelled aperture. Middle: Low-pass filtered correlate: *SLRC*. Bottom: High-pass filtered correlate: *FSRC*.

In Figure 14 the results for pathological simulated voice may be seen. The top template is similar to the *glottal aperture* in Figure 2, whereas the middle one is the low-frequency component of the *glottal aperture*, and the bottom one is the estimation of the *MWC*, which also resembles the one in Figure 11. The results for pathological simulated voice show the presence of *MWC* due to the non-linear effects in the model: the clipping in cord excursion, the interaction between cord position and glottal pressure difference, and the bouncing of one cord against the other. These non-linearities transfer energy from low to high frequencies which appear as a superimposed vibration on the *glottal aperture*. The plot in Figure 15 establishes a comparison among the behavior of the traces used in the study (normal vs pathological, recorded vs simulated), showing the dispersion of the *HNR* ($r_p$) for the estimation windows considered. It may be concluded that the ratio $r_p$ for normal voice is larger than for pathological voice. The results for normal voice from actual recordings and from simulations compare within the same ranges. As simulation results may be adjusted using the model settings, a hint on which parameters may be associated with normal and pathological voice may be obtained through model parameter adaptation: i.e., the comparison between average levels could give an estimation of the possible degree of cord stiffness in a real case under exploration.

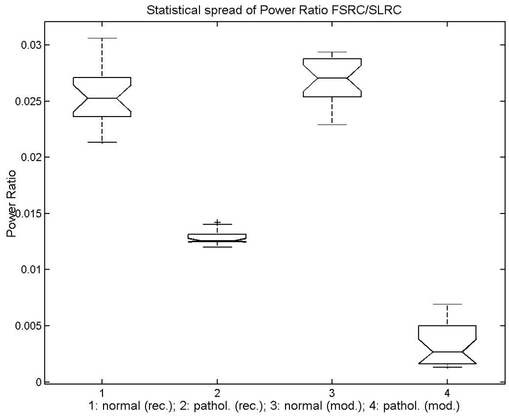

**Figure 15.** Power ratio between *FSRC* and *SRLC* for normal and pathological voice from recordings (rec.) and simulations (mod.).

Results from simulated pathological voice show lower values on the average than recorded pathological, this fact being predictable, as the stiffness induced was the maximum possible, in contrast with pathological recordings. But the most interesting result is that the confidence intervals for pathological and non-pathological traces do not overlap, showing considerable separation intervals, concluding that this parameter (*HNR* between *SLRC* and *FSRC*) could be a good correlate for cord stiffness pathology detection.

## 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.

## REFERENCES

[1] Alku, P., Vilkman, E., "Estimation of the glottal pulseform 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 (to appear).

[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 body-cover model of the vocal folds”, *J. Acoust. Soc. Am.*, Vol. 97, 1995, pp. 1249–1260.