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 traveling 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 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:

\[
\begin{align*}
& f_{i,j} \cdot v_{i,j} - v_{i,j} \cdot R_{i,j} - M_{i,j} \cdot \frac{dv_{i,j}}{dt} - K_{i,j} \cdot \int_{-\infty}^{\infty} v_{i,j} \, dt - \\
& - K_{12l,r,j} \cdot \int_{-\infty}^{\infty} (v_{i,j} - v_{-i,j}) \, dt = 0
\end{align*}
\]

where \( i \in \{1,2\} \) determines the subglottal (1) or the supraglottal (2) cords and \( j \in \{l,r\} \) distinguishes left from right cords, \( f_{i,j} \) 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 \( p_l - p_r \) (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:

\[
u_g = \left( p_l - p_r \right) \frac{L}{\rho c} \int_{-\infty}^{\infty} \left( v_{1l} - v_{1r} \right) \, dt \int_{-\infty}^{\infty} \left( v_{2l} - v_{2r} \right) \, dt
\]

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 \( M_{1,j} \) and \( M_{3,j} \) 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.](image)

![Figure 2. Top: Glottal aperture for modeled cord stiffness. Middle: Derivative of the glottal aperture.](image)

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 $K_{7,2}$ 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.

3. ESTIMATING THE GLOTTAL APERTURE

The reconstruction of the MWC from the voice trace is based in inverting ([4], [3], [11]) the well-known voice production model given in for instance in [4], pp. 193. The voice trace $s(n)$ may be seen as the output of a generation model $F(z)$ excited by a train of delta pulses, its output being modelled by the vocal tract transfer function $F_r(z)$ to yield voice at the lips $s_l$ which is radiated as $s$, where $r = e^{j\omega t}R(z)$ is the radiation model and $f_r$ and $f_g$ are the glottal and vocal tract impulse responses:

$$s = \left( f_g * f_r \right) * r = / f_g * f_r / * r = s_l * r \quad (3)$$

This model will be inverted to reconstruct the glottal aperture $u = \delta f_f$ from the voice trace $s$ by removing the radiation effects to get the radiation-compensated voice $s_l$. A first estimation of the Inverse Glottal Impulse Response $h_g$ may be used to reconstruct the de-glottalized voice $s_g$ from which a first estimation of the Inverse Vocal Tract Impulse Response $h_v$ may be derived, which may be used to remove the influence of the vocal tract from the radiation-compensated voice $s_l$ by direct convolution producing a first estimation of the glottal pulse $u_g$ as summarized in Figure 4.

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

The described procedure has been applied to a trace of non-pathological voice corresponding to the vowel $/\text{a}4/$, 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 DF1 low-pass filtering. The first process is explained in Figure 6.

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

The described procedure has been applied to a trace of non-pathological voice corresponding to the vowel $/\text{a}4/$, 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 DF1 low-pass filtering. The first process is explained in Figure 6.

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

The described procedure has been applied to a trace of non-pathological voice corresponding to the vowel $/\text{a}4/$, 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 DF1 low-pass filtering. The first process is explained in Figure 6.

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

The described procedure has been applied to a trace of non-pathological voice corresponding to the vowel $/\text{a}4/$, 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 DF1 low-pass filtering. The first process is explained in Figure 6.

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

The described procedure has been applied to a trace of non-pathological voice corresponding to the vowel $/\text{a}4/$, 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 DF1 low-pass filtering. The first process is explained in Figure 6.
This signal can be now low-pass filtered using spectral truncation in the frequency domain by means of the DFT:

\[ U_{\text{gl}}(m) = W_{\text{lp}}(m) \sum_{k=0}^{N_k-1} u_{\text{gl},k} e^{-j\frac{2\pi k m}{N_k}} \]  

(7)

\[ u_{\text{gl},k} = \left| \frac{1}{N_k} \sum_{m=0}^{N_k-1} U_{\text{gl}}(m) e^{-j\frac{2\pi k m}{N_k}} \right| \]

(8)

\[ u_{\text{gh},k} = u_{\text{gl},k} - u_{\text{gh},k} \]

(9)

where \( W_{\text{lp}}(m) \) 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 \( u_{\text{gl},k} \) is obtained by inverse DFT and rectification (8), and the high-frequency trace \( u_{\text{gh},k} \) 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 \( u_{\text{lf}} \). Middle: Low-pass filtered glottal aperture \( u_{\text{gl}} \) (SLRC). Bottom: High-pass filtered glottal aperture \( u_{\text{gh}} \) (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:

\[ r_{\text{pk}} = \sum_{n=0}^{N_k} \frac{\left| u_{\text{gh},n}(n) \right|^2}{\sum_{n=0}^{N_k} \left| u_{\text{gl},n}(n) \right|^2} \]

(10)

where \( w_k \) is the \( k \)-th period-adjusted window used in the evaluation of \( r_{\text{pk}} \), \( u_{\text{gl},n} \) is the levelled glottal aperture, and \( u_{\text{gh},n} \) 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 hypothesis a new set of traces was produced from the vocal cord model with the following settings for normal voice: \( M_{l,r1}=0.2 \text{ g}, M_{l,r2}=0.02 \text{ g}, K_{l,r1}=40,000 \text{ dyn/cm}, K_{l,r2}=100,000 \text{ dyn/cm}, K_{l,r3}=30,000 \text{ dyn/cm}, \) and the same parameter values for pathological voice with the exception: \( M_{l,r2}=0 \text{ g}. \) Normal voice traces from simulations are given in Figure 13.

![Figure 13](image-url)

**Figure 13.** Traces from simulations (normal case): Top: Levelled signal. 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 also resembles the one in Figure 11. The results for 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.

![Figure 14](image-url)

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

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**


