Evidence of Glottal Source Spectral Features found in Vocal Fold Dynamics

Gómez, P., Godino, J. I., Álvarez, A., 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

Pathology detection methods in vocal cords using voice may benefit from the detection of the time and frequency patterns of the *mucosal wave correlate*. In the present work a description of the vocal folds is given in terms of the most relevant features related to vocal fold pathology in reference to the body-cover models of 2- and k-masses. A description of the *mucosal wave correlate* as the residual *small-amplitude vibration* when removing the body influence in the *glottal aperture* is given. The main features of the *mucosal wave correlate* will be commented. Finally, a parameterization method based on this study and its possible applications to speech pathology detection will be discussed.

1. INTRODUCTION

It is well known from literature that certain pathologies of the vocal folds influence the power spectrum of voice to the point that certain frequency-domain characteristics in the spectrum are relevant to the determination of the presence of pathology in the vocal fold system. In this sense, the use of parameters as the acoustic features H1-A1, H1-A3, A1-A3 and H1-H2 is well documented in [7]. Some of the inconveniences found in using these parameters lay in the influence of other parts of the voice production system which may not be relevant in the vocal pathology detection process, and which obscure or hide somehow the results, as is the case of the vocal tract. In this sense the first order vibration mode of the vocal folds is influenced mainly by the presence of the body masses of vocal cords, and make the observation of fine second-order effects -as the *mucosal wave*- difficult. This fact is especially relevant, as the presence of *mucosal wave* is strongly altered by dynamic changes in the vocal folds [8] due to pathology. In the present study it will be shown that the *mucosal wave correlate* present in the glottal source is an important clue to spectral features which are closely related to second and higher order dynamics of the vocal folds, and from which important knowledge may be derived oriented to pathology detection and classification.

2. VOCAL CORD STRUCTURE AND K-MASS MODELS

The physiological structure of the vocal cord parts involved in the glottal closure are seen in Figure 1.a, in which the massive part of the cord, composed mainly of muscle is described as "the body", and the epithelial and tissular envelope of the cord (*lamina propria*) is known as "the cover". The modelling of the mentioned structure as a set of lumped masses joined by elastic springs is shown in Figure 1.b. Each mass will be linked to the body mass and to two neighbour ones, except the end masses. The main masses are associated to the body and to the supraglottal and subglottal ridges, although other distributions are possible. In general, the largest masses would be associated with the supraglottal and subglottal lips or ridges controlling the glottal closure. Masses can move both horizontally or vertically, and their relative movement is governed by a complicated pattern denoted as *mucosal wave*, which is in essence a kind of a *travelling wave* propagating along the cover, and inducing nonlinear effects when both cords are close enough due to cord collision [1].

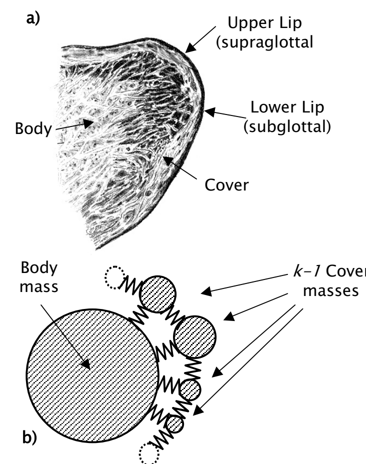

Figure 1. a) Body-cover structure of a vocal cord seen in cross-section. b) k-mass equivalent model.

For the purposes of this study, it will be considered that they are allowed to move only horizontally (in the sense of approaching or separating to each other cord). The *mucosal wave correlate* (*MWC*) will appear as a superimposed waveform on *Glottal Aperture (GA)*, as shown in Figure 2, in which a dominant movement of large amplitude characterised by an arch-like pattern is supporting a small-amplitude waveform [10], resulting from the movement of the vocal cords known as *mucosal wave*.

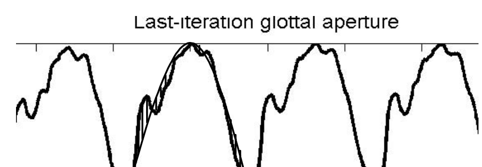

Figure 2. Relation between *large- and small-amplitude components (LAC* and *SAC)* and the *mucosal wave correlate*.

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, as signalled in the figure by the dashed vertical lines. As it has been commented before, the lumped masses associated to the cover are linked together and to the body. For the purposes of our study, a 2-mass model derived from the well known of Ishizaka and Flanagan will be used [6], assuming that masses are associated in pairs relative to the reference mass, either this being the walls of the larynx or the body mass. In this way the k-mass model would be reduced to several 2-mass models in a *divide and conquer* strategy. For this study a version of the *vocal cord 2-mass model* as given in [6] has been implemented in MATLAB® [3], its main features being: 2-mass asymmetric modelling, non-linear coupling between mass movement and *GA*, 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 section of the cord)* M_{il} and M_{jl} (left cord), M_{ir} and M_{jr} (right cord), the *elastic parameters* K_{il} and K_{jl} (relative to reference) and K_{ijl} (intercoupling), and their respective ones for the right cord: K_{ir}, K_{jr} and K_{ijr}. 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

$$f_{xps} - v_{ps}R_{ps} - M_{ps}\frac{dv_{ps}}{dt} - K_{ps}\int_{-\infty}^{t} v_{ps}dt - \qquad (1)$$
$$- K_{pij}\int_{-\infty}^{t}(v_{pi} - v_{pj})dt = 0$$

where $p \in \{1,k\}$ determines the specific mass (i) or (j) and $s \in \{l,r\}$ distinguishes left from right cords, f_{xps} 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 (the excitation), and v_{ps} is the corresponding mass speed along the axis x (the response).

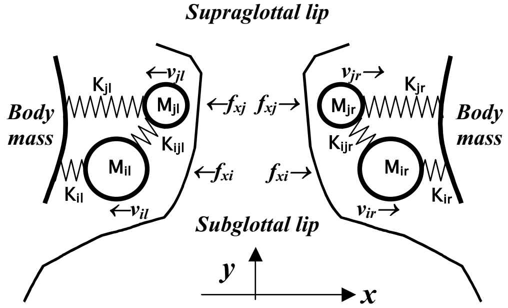

Figure 3. Schematic structure of one section of the cover modeled as a two-mass system referred to the body mass.

This model is a simplification of more elaborate ones [9], and will be used to reproduce the vibration features of a pair of cover masses linked to a bulky body mass, the results to be extended to a system of k-masses of interest for the present study. The structure of the model may be seen in Figure 3.

3. REMOVING THE EFFECTS OF THE BODY MASS

The reconstruction of the *MWC* from the voice trace is based in inverting the well-known voice production model given in [2], pp. 193. The algorithmic details of the estimation procedure for the *Glottal Aperture (GA)* $\mathbf{u_g}$ are given in [5]. This procedure has been applied to a trace of normophonic voice corresponding to the vowel **/a/**, of which a segment of 0.05 sec. of duration is shown in Figure 4. Several techniques have been used to remove the *large-amplitude component (LAC)* and produce an estimate of the *MWC*, as mean-, low pass- and cepstral filtering [5], showing good results when the *GA* minima are not too sharp, otherwise the residual component of the *LAC* near the minima is large and it distorts the *MWC* estimate.

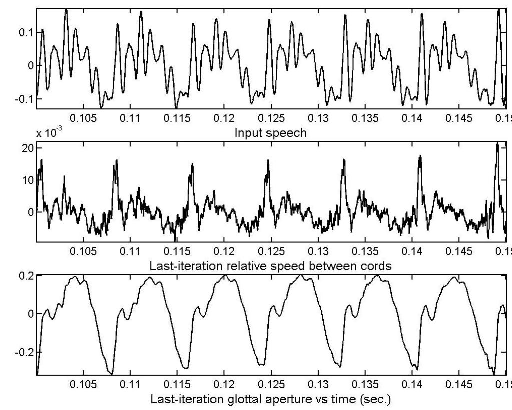

Figure 4. Normophonic voice. Top: Input voice. Middle: Differential GA. Bottom: GA.

The technique proposed here is based on a period-by-period subtraction of the slow-moving baseline on which the minima of the *GA* relies, and on DFT low-pass filtering. For such the corresponding increment $\Delta u_{g,n}$ at time instant n will be subtracted from the *GA* $u_{g,n}$

$$u_{gf,n} = u_{g,n} - \Delta u_{g,n} = = u_{g,n} - \frac{u_{g\,min,k} - u_{g\,min,k-1}}{m_k - m_{k-1}}(n - m_{k-1}) \qquad (2)$$

$$u_{gm,n} = u_{gf,n} - \min\{u_{gf,n}\} \qquad (3)$$

$$u_{gu,n} = (-1)^k u_{gm}(n \in w_k) \qquad (4)$$

where $u_{gmin,k}$ and $u_{gmin,k-1}$ are the respective values of the peaks at positions m_k and m_{k-1}, and w_k is the *k-th cycle window*. By subtracting the minima of *(2)* and sign reversal of each alternate cycle-window w_k *(4)* the *unfolded GA* $u_{gu,n}$ is obtained (see Figure 5), which could be seen as the average excursion described by one of the cords if left vibrating freely (no opposite cord).

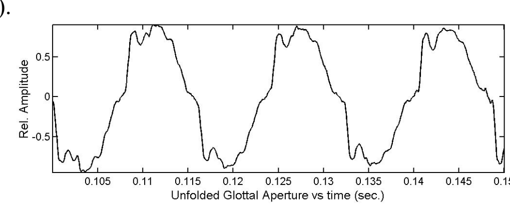

Figure 5. Unfolded GA $u_{gu,n}$.

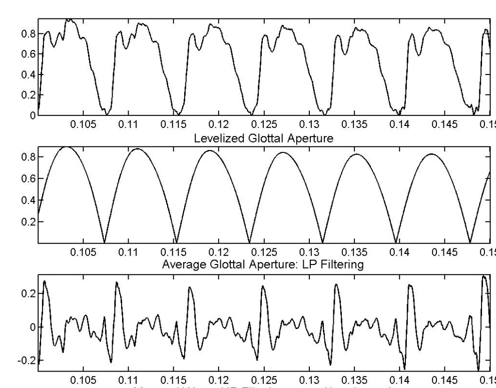

Figure 6. Normophonic voice. Top: Levelled GA: $u_{gf,n}$. Mid: Low-pass filtered GA: $u_{gl,n}$. Bottom: MWC: $u_{gh,n}$.

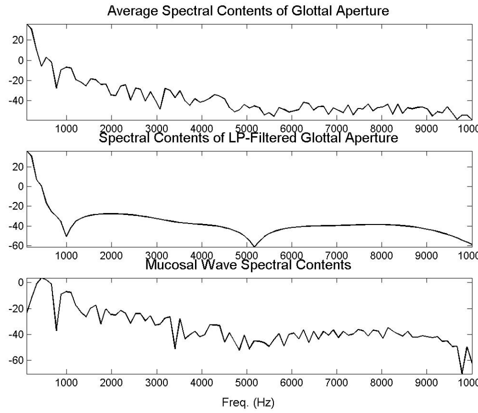

Figure 7. Typical power spectral densities of Levelled GA (top), Low-pass filtered GA (midd.) and High-pass filtered GA (mucosal wave correlate) for normal voice.

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

$$U_{gb}(m) = W_{lp}(m) \sum_{n \in w_k} u_{gu,n} e^{-jm\frac{2\pi}{N_k}n} \qquad (5)$$

$$u_{gl,n} = \left| u_{gb,n} \right| = \left| \frac{1}{N_k} \sum_{m=0}^{N_k - 1} U_{gb}(m) e^{jn\frac{2\pi}{N_k}m} \right| \qquad (6)$$

$$u_{gh,n} = u_{gb,n} - u_{gl,n} \qquad (7)$$

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 (6), and the high-frequency trace $u_{gh,n}$ by subtraction (7), the resulting traces shown in Figure 6. Their associated power spectral densities are given in Figure 7.

4. *MWC* FEATURES FROM VOCAL FOLD DYNAMICS

It may be interesting to consider if the proposed extraction method yields results which are in agreement with vocal cord dynamics, assessing if what is called *mucosal wave correlate* is really related to the physical essence of the *mucosal wave*.

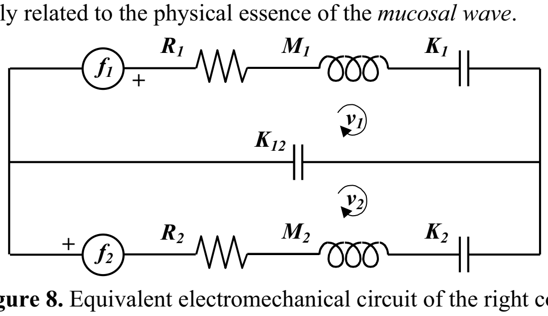

Figure 8. Equivalent electromechanical circuit of the right cord in the two-mass model.

For such the spectral power density of *the mucosal wave correlate* obtained from a specific case of normophonic voice as given in Figure 7 (bottom) –also Figure 9.b- will be compared 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_i and f_j acting on both masses of the same cord (as for instance M_{ir} and M_{jr}) with the observable mass velocities v_i and v_j in the electromechanical equivalent to the one-cord dynamical system given in Figure 8, derived from the respective dynamical equations given in (1). Details of this study are given in full length in [4]. When plotting the *cord trans-admittance* relating f_i and v_i in the frequency domain, using *ad hoc* values for the biomechanical parameters of the model, a curve as the one given in Figure 9.a may be found to match the case given in Figure 9.b, showing a notch between two maxima for frequencies below 1000 Hz which largelly resemble the general pattern of the *MWC* power spectral density.

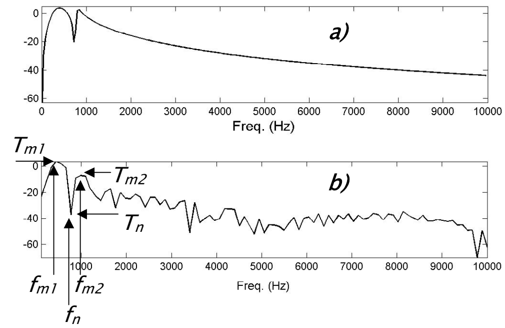

Figure 9. a) Square modulus of the trans-admittance between f_i and v_i in the frequency domain (permanent regime) assuming f_j=0. b) Power spectral density of the mucosal wave correlate for a typical voice frame (normophonic).

It will be now explored if the singularity found in the *mucosal wave correlate* is a normal feature present in normophonic voice. For such the traces produced by *16 normophonic speakers* were processed, half of them males and half females, uttering the vowel */a/* for about *2 sec*. The utterances were clipped for *0.2 sec*. around their middle part (to avoid onset and decay effects), and the singularity parameters were evaluated over an average of the number of cycles present in the trace, in the range *18-40*, depending on each case. The results are shown in Table 1.

Subj.	Sex	Val. 1st max. (dB)	Freq. (Hz)	Val. min. (dB)	Freq. (Hz)	Val. 2nd. max. (dB)	Freq. (Hz)
#6.4	M	3.62	439	-36.95	769	-6.97	989
#6.5	M	12.38	467	-10.65	747	1.76	841
#6.7	M	8.04	344	-3.01	574	6.41	804
#6.8	M	-2.14	370	-12.99	493	-7.44	617
#6.B	M	15.42	336	-15.58	504	8.36	588
#7.1	M	2.41	357	-27.17	833	-11.16	1071
#6.E	M	11.85	434	-12.28	869	-2.15	978
#6.F	M	5.43	404	-20.48	707	-12.33	909
#0.4	F	-1,76	416	-18.92	833	-15.09	972
#0.7	F	-1.72	638	-24.72	1063	-21.32	1276
#1.2	F	3.42	769	-20.05	1346	-14.87	1730
#1.3	F	-7.10	612	-30.62	1020	-9.65	1632
#2.6	F	4.50	689	-19.11	2069	-12.16	2413
#2.9	F	2.08	888	-23.90	2222	-18.67	2444
#2.A	F	-7.63	666	-24.93	1111	-14.93	1333
#2.B	F	-6.47	784	-24.52	1176	-17.75	1568

Table 1. Results after detecting the singularities found in the *mucosal wave correlate* for *16* normophonic subjects.

The first conclusion from the table is that female voices show positions for the singularities at higher frequencies than male voices, as should be expected. Besides, male voices show a sharper notch than female voices. To help in comparing the absolute values given in Table 1, a normalized set of values is given in Table 2. From left to right the third column gives the depth of the notch relative to the first maximum in dB ($T_{rn}=T_n-T_{m1}$). The fourth one gives the position of the notch relative to the position of the first maximum ($f_{rn}=f_r/f_{m1}$). The fifth one gives the height of the second maximum relative to the first one. The sixth one gives the position of the second maximum relative to the first one ($f_{r21}=f_{m2}/f_{m1}$). The seventh column gives the notch slenderness factor, which is defined as

$$N_{sf}=\frac{T_{rn}}{f_{r21}-1} \quad (8)$$

The eighth column (right-most) gives the normalized position of the notch relative to the positions of the two maxima, defined as

$$p_{rn}=2\frac{f_{rn}}{f_{r21}+1} \quad (9)$$

It may be seen in this case that although the slenderness of the notch is rather variable between subjects (being larger in general for males than for females), the relative position of the notch to the central position between the maxima is more regular for most of the subjects.

Subj.	Sex	Notch Rel. Depth	Notch Rel. Posit.	2nd Max. Rel. Height	2nd Max. Rel. Posit.	Notch Slend. Factor	Notch Norm. Posit.
#6.4	M	-40,57	1,75	-10,59	2,25	32,38	1,08
#6.5	M	-23,03	1,60	-10,62	1,80	28,76	1,14
#6.7	M	-11,05	1,67	-1,63	2,34	8,26	1,00
#6.8	M	-10,85	1,33	-5,30	1,67	16,25	1,00
#6.B	M	-31,00	1,50	-7,06	1,75	41,33	1,09
#7.1	M	-29,58	2,33	-13,57	3,00	14,79	1,17
#6.E	M	-24,13	2,00	-14,00	2,25	19,25	1,23
#6.F	M	-25,91	1,75	-17,76	2,25	20,73	1,08
#0.4	F	-17,16	2,00	-13,33	2,34	12,84	1,20
#0.7	F	-23,00	1,67	-19,60	2,00	23,00	1,11
#1.2	F	-23,47	1,75	-18,29	2,25	18,78	1,08
#1.3	F	-23,52	1,67	-2,55	2,67	14,11	0,91
#2.6	F	-23,61	3,00	-16,66	3,50	9,44	1,33
#2.9	F	-25,98	2,50	-20,75	2,75	14,83	1,33
#2.A	F	-17,30	1,67	-7,30	2,00	17,27	1,11
#2.B	F	-18,05	1,50	-11,28	2,00	18,05	1,00

Table 2. Normalized results for peak slenderness and position.

The variability found is due to the specific phonation characteristics and style of each speaker, and does not put an inconvenience to the study, as the main challenge is how to use the positions and values of the singularities to determine the values of the biomechanical parameters of the associated *2-mass model*, this being an objective left for future study.

5. CONCLUSIONS

Through the present paper it has been shown that specific characteristics present in the power spectral density of the *mucosal wave correlate* can be due to the singularities present in the transfer function of the *vocal cord trans-admittance* relating the force exerted on one of the cord masses and the velocity of the same mass in the frequency domain. These singularities appear as two maxima in the transfer function and a minimum between them. As there is a clear connection between the singularity positions and size and the biomechanical parameters associated to the *2-mass model* supporting the general form of the *trans-admittance*, the biomechanical parameters could be estimated from the direct measurement of the positions, values and bandwidths associated to the singularities, thus opening the door to non-invasive characterization of the vocal cord system with applications in pathology screening. This behavior may be found in all 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 masses and springs relative to the body, and the notch position and size is due to the intercoupling between masses (K_{ij}). Curve fitting may grant optimal detection of biomechanical parameters best approaching each case under study. For systems which show several notches encapsulated between maxima, a model may be established incorporating a section of two 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.

6. ACKNOWLEDGMENTS

This research is funded by the following grants 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] Berry, D. A., Doellinger, M., Berke, G. S., “Medial Surface Dynamics of the Vocal Fold During Phonation”, *Proc. of the International Conference on Voice Physiology and Biomechanics*, Marseille, France, August 18-20, 2004, pag. 5.

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

[3] 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.

[4] 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).

[5] 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.

[6] 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.

[7] Kuo, J., Holmberg, E. B., Hillman, R. E., “Discriminating Speakers with Vocal Nodules Using Aerodynamic and Acoustic Features”, *Proc. of the ICASSP’99*, Vol. 1, 15-19 March 1999, pp. 77-80.

[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] Titze, I. R., “The physics of small-amplitude oscillation of the vocal folds”, *J. Acoust. Soc. Am.,* Vol. 83, 1988, pp. 1533-1552.