

Wavelet Description of the Glottal Gap

Pedro Gómez-Vilda, Víctor Nieto-Lluis, Victoria Rodellar-Biarge, Rafael Martínez-Olalla, Cristina Muñoz-Mulas, Agustín Álvarez-Marquina, Luis M. Mazaira-Fernández
Neuromorphic Speech Processing Lab
Centro de Tecnología Biomédica, Universidad Politécnica de Madrid
28223 Pozuelo de Alarcón, Madrid, Spain
e-mail: pedro@fi.upm.es

Bartolomé Scola-Yurrita, Carlos Ramírez-Calvo,
Daniel Poletti-Serafini

Ear, Neck and Throat (ENT) Services
Hospital Universitario Gregorio Marañón
C/ Maiquez, 7
28009 Madrid, Spain

Abstract— The Glottal Source correlates reconstructed from the phonated parts of voice may render interesting information with applicability in different fields. One of them is defective closure (gap) detection. Through the paper the background to explain the physical foundations of defective gap are reviewed. A possible method to estimate defective gap is also presented based on a Wavelet Description of the Glottal Source. The method is validated using results from the analysis of a gender-balanced speaker's database. Normative values for the different parameters estimated are given. A set of study cases with deficient glottal closure is presented and discussed.

I. INTRODUCTION

The production of voice is due to the vibration induced in the Vocal Folds by transglottal pressure. The Vocal Folds are part of the Larynx, composed by a complex system of cartilages, tiny muscles and connective tissues. The function of the vocal folds may be seen summarized in Figure 1. The phonation cycle starts when the vocal folds in inspiration (a) are brought together (adduction) by laryngeal muscles (b). This closes glottis (contact phase). The vocal folds are separated (abduction) briefly (c) by trans-glottal pressure (open phase).

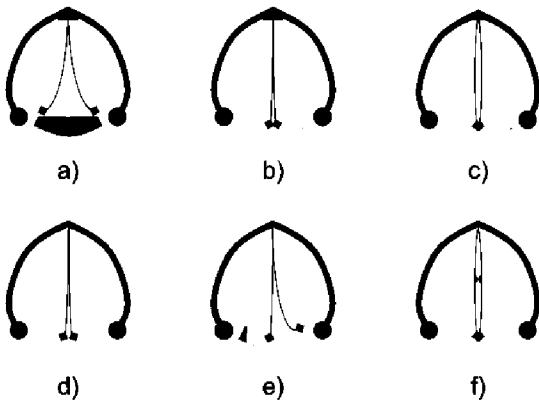


Figure 1 Glottal gap as part of the phonation cycle. The gap may be seen as the average opening between the vocal folds (vertical bands). a) Glottis open. b) Glottis closed. c) Glottal gap during phonation. d) Defective Permanent Gap (DPG). e) Defective gap caused by lateral paresis. f) Example of Defective Contact Gap (DCG) caused by lesion (bilateral nodules, in red).

Elastic and dynamic forces on the vocal folds, bring them together again, and the cycle is repeated around 100 times per

second in male voice and about twofold that in female voice. As the vocal folds are three dimensional bands, the phonation cycle described is much more complex. The Glottal Gap may be roughly described as the average light or aperture seen by the air flow at any time (the transversal surface seen through both vocal folds [1]). The phonation cycle is composed of a closed phase (where gap is almost null) and an open phase, where the gap should start from zero to reach a maximum and to zero again at the end of the cycle. But this ideal situation is rarely attained. It may be that for anatomical reasons, or for the lack of tension in the muscles (asthenic voice, whispery voice, etc.) the closure is not complete on the side of the arytenoid horns (d), leaving a small gap even during the closed phase. This situation is referred to as Defective Permanent Gap (DPG). Sometimes one of the vocal folds may not fully contact (e). In other cases, some organic lesion on the vocal folds may prevent complete closure (f), as in the case of polyps, nodules, cysts, etc, altering the normal vibration of the vocal folds (mucosal wave, see [2]). In such cases, air may escape when it was supposed to be stopped. This situation is known as Defective Contact Gap (DCG). The relation between gap defects and voice pathology is well known [1]-[3]. The purpose of the present paper is to give estimations of the defective gaps described above from a wavelet description of the glottal source after inverting the voice signal [4], [5]. Wavelet methods have been used before in assessing voice pathology [6] although not directly on the glottal source. In Section II a simplified model of the glottal gap is presented. Simulations using the gap-generating model are given in Section III to justify the use of wavelet models in describing the nature of defective gaps as wavelets of the main glottal cycle. Section IV is devoted to process a database of normophonic subjects (male and female) to obtain normative descriptions of the singularity points and the values of the DPG and DCG in a general population. Descriptive statistics of these estimates will be given, and specific dysphonic cases are studied. Finally Section V presents conclusions and describes possible future research.

II. GAP SIMPLIFIED MODEL

The purpose of the present section is to define a model relating the pattern of the average effective gap (glottal aperture through glottis) and the glottal source time profile. This is done by simplifying the dynamic flow equations in the glottis relating air flow and pressure [7]. The following assumptions hold: plane wave propagation, single tube straightened lossless vocal tract, no nasalization, the vocal

folds represented as a point-like switch with a cross-effective section (gap) described by a non-linear conductive parameter $G_g(t)$, and full cylindrical symmetry. The lungs and trachea will be seen as a flow-injecting compliant system. This simplified gap model may be seen as a reduction of the well-known Klatt model [8], as presented in Figure 2. The dynamic variables are the lung injected flow u_l , the sub- and supra-glottal pressures p_l and p_g , and the glottal flow u_g . The dynamic pressure at the lips will be zero (atmospheric reference: $p_0=0$). The excitation is provided by the glottal conductance $G_g(t)$. The lung-trachea system is represented by the compliance C_l . The simplified vocal tract is represented by the inertance L_t . The electromechanical model will then be

$$\begin{aligned} u_l(t) &= C_l \frac{\partial p_l}{\partial t} + u_g \\ p_g &= L_t \frac{\partial u_g}{\partial t} \\ u_g &= G_g(p_l - p_g) \end{aligned} \quad (1)$$

These equations may be solved numerically in the time domain given a specific gap function G_g to reproduce Liljencrants-Fant (L-F) pattern of the glottal source [8]. It shows a DCG during the closed phase (small opening with amplitude 0.1 times the main opening between 2 and 3 ms), and a DPG (a permanent opening 0.15 times the main opening).

III. WAVELET MODELING OF THE GLOTTAL GAP

A gap function $G_g(t)$ as the one in Figure 3 (a) is defined to simulate the phonation cycle with a permanent gap (DPG) of 0.1 times the glottal opening. As a contrast a new glottal gap function $G_g(t)$ including a DCG equivalent to 0.2 times the glottal opening is simulated in Figure 3 (b).

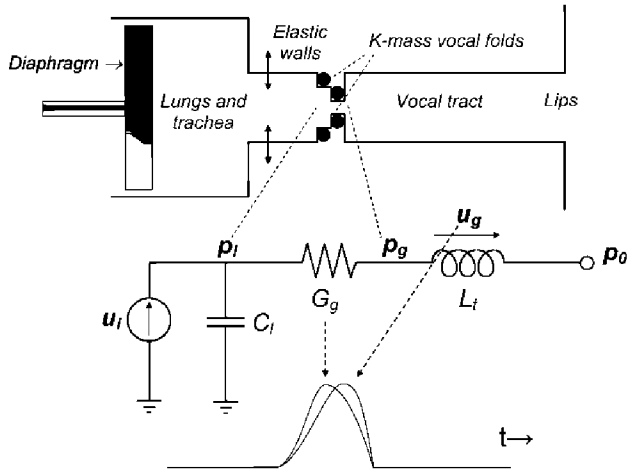


Figure 2 Top: Simplified gap model. Bottom: electromechanical equivalent.

The values of the model parameters are: $u_l=0.005 \text{ m}^3\text{s}^{-1}$; $L_t=0.003 \text{ Pa}\cdot\text{m}^{-3}\text{s}^2$; $C_p=0.001 \text{ Pa}^{-1}\text{m}^3$. The glottal source shows a pattern (encircled in the cartouche) which is a scale reproduction in time and amplitude of the L-F pattern, induced by the DCG. The effect of the DPG is a reduction of the slope

in the return phase following the closing instant (from 0 to 0.1 ms in Figure 3 upper template, blue line). The effect of the DCG can be considered as a wavelet reproduction of the main phonation pattern. Therefore wavelet theory may be used to describe and estimate defective contact gaps. A description of the glottal source $p_g(t)$ in terms of Haar's wavelet [9] will be proposed

$$\begin{aligned} \psi_{00}^g(t) &= \begin{cases} 1; & 0 \leq t < \frac{t_c}{2} \\ -1; & \frac{t_c}{2} \leq t < t_c \end{cases}; \\ \psi_{jk}^g(t) &= 2^{-j/2} \psi_{00}^g(2^{-j}t - k\tau); \end{aligned} \quad (2)$$

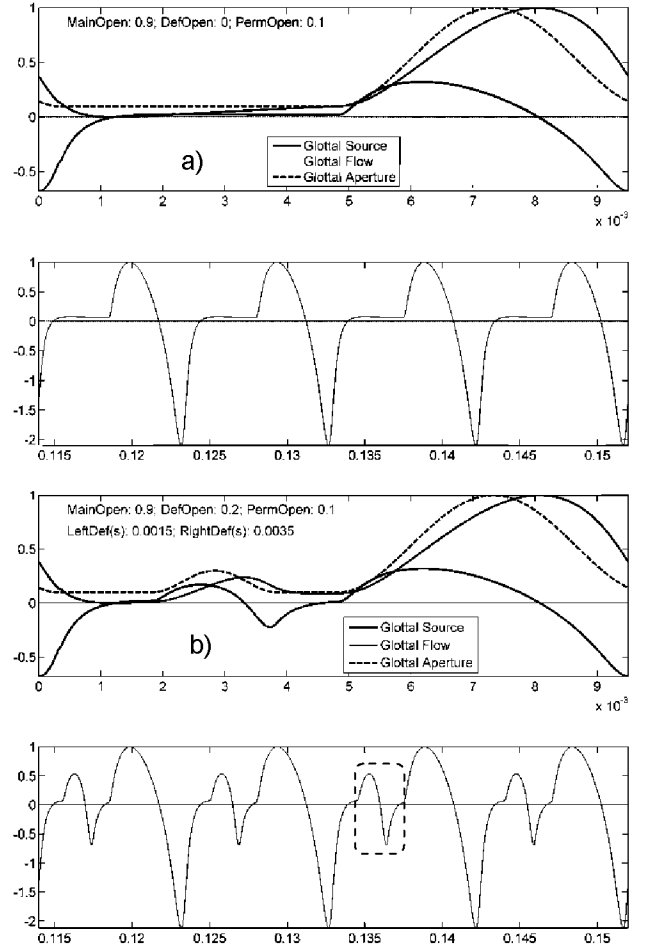


Figure 3 a) Top: L-F pattern of the glottal source (blue thick line) under permanent gap conditions (red dash line), and glottal flow (green light line) correspondent to a glottal cycle 9.5 ms long. Bottom: four glottal phonation cycles. b) Top: glottal source when a small contact defect is present (hump in the dot-red plot between 2 and 4 ms), and the associate glottal flow (green light line). Bottom: four glottal phonation cycles. Horizontal axis given in time (s).

where t_c is the closing instant, j and k are dilation and position indices, and τ is the sampling interval. The wavelet detail at scale j is defined as

$$p_{dj}(\xi = k\tau) = \int_{-\infty}^{\infty} \psi_{jk}^s(t) p_g(t) dt. \quad (3)$$

The wavelet analysis of the two glottal source examples in Figure 3 is given in Figure 4. The instants of opening t_O , maximum t_M , closure t_C and recovery t_R may be inferred from the minima in the first detail (a). Similarly the gap opening t_{Og} , maximum t_{Mg} and closure t_{Cg} are estimated from the first detail of (b). The relevant points for our study are the gap opening instant (t_{Og} , start of the DCG around 2 ms in the upper template of Figure 3), and the true opening instant (t_O , around 5 ms). The glottal source maxima for the first and second opening intervals are also relevant. These points may be estimated from the wavelet first order detail, as follows

$$\begin{aligned} t_{Og} &= \max \left\{ \arg \min_{0 \leq \xi < t_{Mg}} \text{rel} \{ p_{d1}(\xi) \} \right\} \\ t_{Mg} &= \max \left\{ \arg \min_{0 \leq \xi < t_O} \text{rel} \{ p_{d1}(\xi) \} \right\} \\ t_O &= \max \left\{ \arg \min_{0 \leq \xi < t_M} \text{rel} \{ p_{d1}(\xi) \} \right\} \\ t_M &= \max \left\{ \arg \min_{0 \leq \xi < t_C} \text{rel} \{ p_{d1}(\xi) \} \right\} \end{aligned} \quad (4)$$

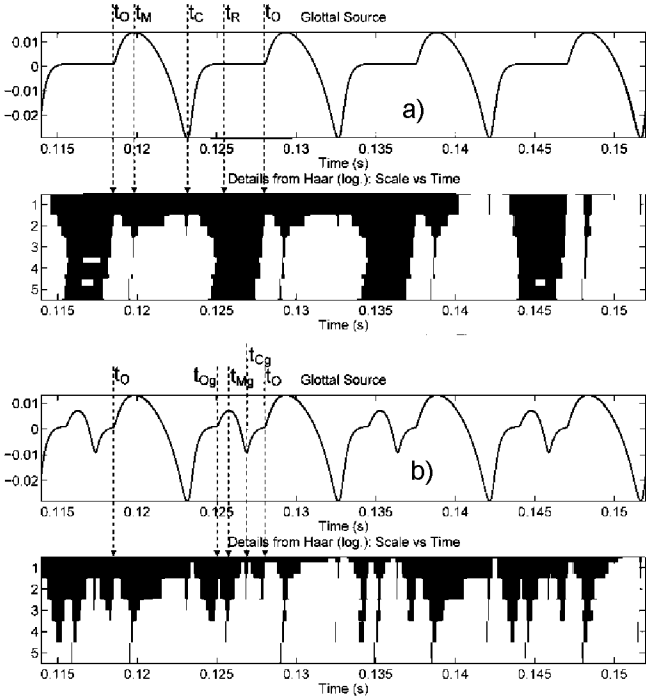


Figure 4 a) Top: synthetic glottal source showing a slight DPG. Bottom: first five details from its Haar wavelet decomposition. b) Top: synthetic glottal source showing a DCG. Bottom: first five details from its Haar wavelet decomposition. Detail plots: position is given in the horizontal axis, scale in the vertical axis.

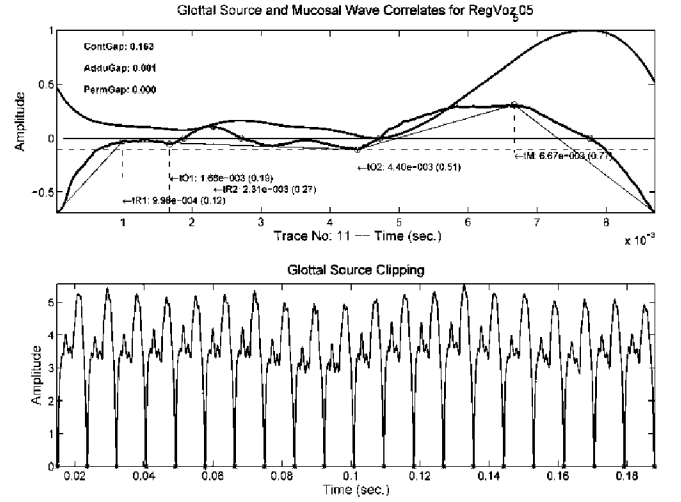


Figure 5 A real case glottal source showing DCG due to bilateral nodules. The expected wavelet behavior is present in the first hump and trough found after t_{O1} (upper template).

The estimation of the DCG is straight forward as the difference between the glottal source amplitudes at t_{Mg} and t_O . The estimation of the DPG is derived from the slope of the glottal source in the interval $0 \leq t < t_{R1}$ (return phase, see Figure 5, upper template). To give an example on the results of this detection methodology Figure 5 shows a real case of a glottal source obtained from a sustained phonation of vowel /a/ by a male subject, from which 200 ms were processed following [10]. This waveform shows a defective closure gap produced by a mild case of bilateral nodules affecting both Vocal Folds (corresponding to grade 1 in Hirano's GRBAS scale evaluating Grade, Roughness, Breathiness, Asthenia and Strain [11]). The wavelet is observable as an over-undershot pattern between $t_{O1}=1.68$ ms and $t_{O2}=4.40$ ms (which stand respectively for t_{Og} and t_O). The purpose of the wavelet description is to detect DCG's and to estimate their extension and amplitude.

IV. RESULTS AND DISCUSSION

A gender-balanced database of sustained /a/ recordings using high-quality headset microphones (modal phonation) from 100 adult subjects (50 male and 50 female) collected at the ENT Services (Ear, Neck and Throat, Hospital Universitario Gregorio Marañón of Madrid, Spain) was processed. All subjects passed endoscopic and subjective evaluation by Hirano's GRBAS scale [12] to disregard organic pathologies. The glottal source $p_g(t)$ was obtained from segments 200 ms long after vocal tract inversion [11]. The wavelet description of the glottal source given in (4) was produced. For validation purposes the gap and true opening times (t_{Og} and t_O) and the DCG and DPG are estimated for each subject in the database. The results for a normative male and female population are given in TABLE I (in quartiles) and Figure 6. The following observations may be deduced from a first analysis of the normative results:

- The false opening time t_{O1} is smaller for the female than for the male group. This is also the case of t_{O2} , as it would be expected [12] but besides, female distribution is bimodal, with its second mode aligned with the male

distribution. A possible hypothesis for this bimodality could be the periodic variability of female voice due to hormonal changes which would produce two groups of t_{O2} estimates, one more in line with male estimates than the other, this being a matter for further study.

- DCG shows that most normophonic speakers present a moderate air escape during the contact phase.
- DPG on its turn is distributed differently in men and women as separate bells. This means that permanent air escape seems to be larger in women than in men.

	t_{Og}	t_o	DCG	DPG
MNQ1	0,2167	0,4483	0,0882	0,1385
MNQ2	0,3202	0,4959	0,1208	0,1696
MNQ3	0,3864	0,5497	0,1534	0,2068
FNQ1	0,1627	0,3099	0,0573	0,3029
FNQ2	0,2031	0,3583	0,0916	0,3501
FNQ3	0,2610	0,5135	0,1571	0,3851

To exemplify the potential use of this methodology in pathology detection 16 patients with pathologies which may produce defective gaps were studied (see TABLE II). The cases selected were possible candidates to present defective contact and/or permanent deficient closure. The contrastive study is carried out relative to the inter-quartile interval for male normophonic distributions ($MNQ1, MNQ3$), and similarly for female normophonic ones ($FNQ1, FNQ3$). The following results derived from the study are the most relevant ones observed:

- In all cases the opening instant t_o was below $MNQ1$ in males. The same was observed in females relative to $FNQ1$ except for case #287. The conclusion derived is that pathologies with deficient closure seem to show a reduced contact phase (time interval between t_c and t_o , during which vocal folds are in contact closing the air passage).
- DCG was out of the interquartile interval in 11 out of 16 cases, either being larger (4 cases) or lower (7 cases) than expected. In general DCG was low in Reinke's Edema and large in pathologies with a high G mark in GRBAS.
- DPG was out of the modal interval in 8 out of 16 cases, most of the times under the inter-quartile distribution (1 case in males, 6 cases in females), possibly due to overpressed (harsh) phonation to compensate air escape.
- The closed phase is shorter in females and less prone to air escapes during the contact, therefore the DCG seems to be larger in males than in females.
- On the other hand the contact is less complete in female voice and a larger DPG is observed in their voices.

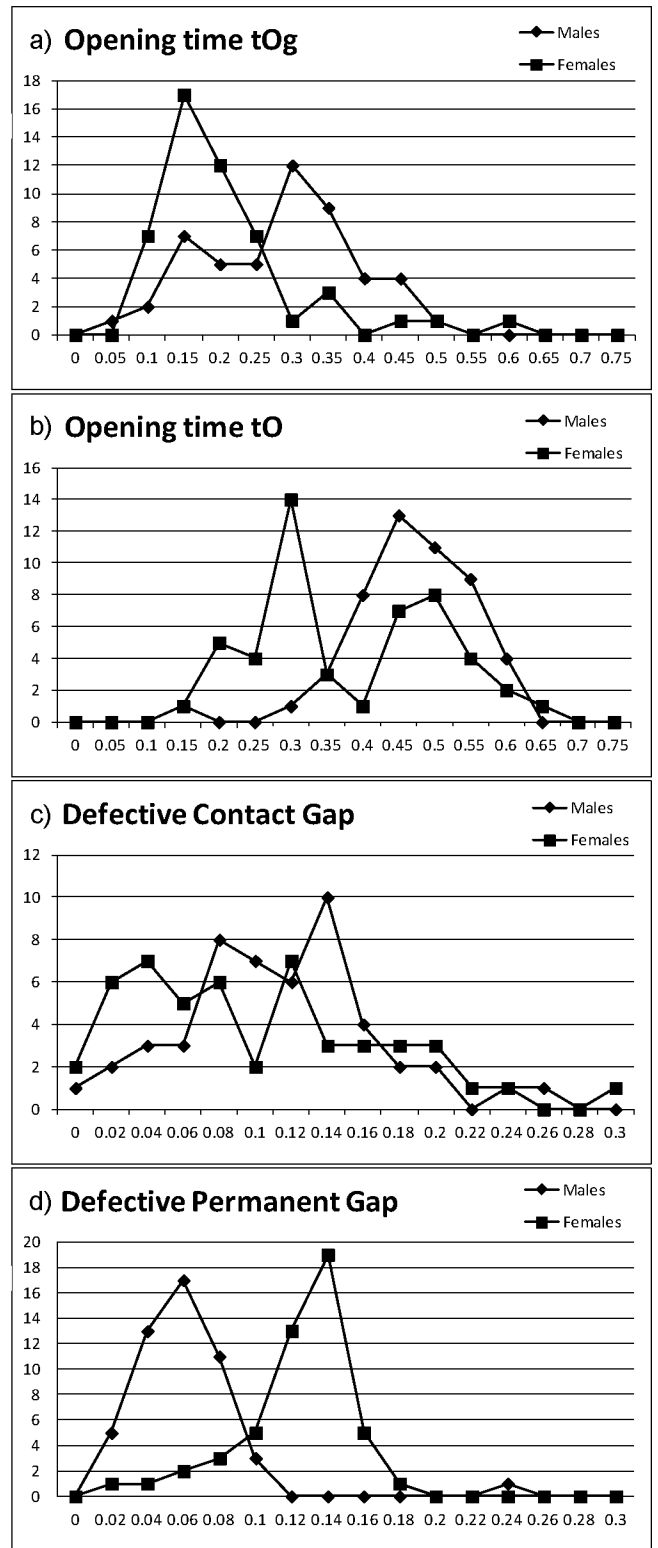


Figure 6 Histograms of the estimated parameters from 50 male and 50 female subjects: a) gap ($t_{O1}=t_{Og}$) opening instant; b) true opening instant ($t_{O2}=t_o$); c) DCG; d) DPG. Ordinates give number of cases, abscisae in a) and b) give time intervals relative to glottal cycle duration, abscisae in c) and d) give relative values of DCG and DPG.

TABLE II RESULTS FOR THE DYSPHONIC CASES STUDIED. AMOUNTS RELATIVE TO GLOTTAL CYCLE DURATION. VF: VOCAL FOLDS

G	Pat.#	Diagnose	t_o	DCG	DPG
M	21	Chronic Laryngitis	0,2745	0,2877	0,0977
M	74	Right VF Paresis	0,1943	0,2603	0,1685
M	216	Functional	0,3873	0,0798	0,2206
M	218	Functional	0,2380	0,1260	0,1257
M	254	Contact Granulomae	0,2605	0,3437	0,1240
M	267	Functional	0,2185	0,1483	0,1931
M	328	Bilateral Nodules	0,3109	0,1383	0,1402
M	438	Left VF Paresis	0,4284	0,0227	0,1995
F	48	Functional	0,1959	0,2670	0,2853
F	69	Functional	0,2292	0,0569	0,2991
F	113	Functional	0,1971	0,0798	0,2623
F	119	Right VF Paresis	0,1998	0,0375	0,3218
F	230	Functional	0,1374	0,0170	0,3266
F	276	Reinke's Edema	0,2124	0,0099	0,1660
F	281	Bilateral Nodules	0,2176	0,0480	0,2092
F	287	Reinke's Edema	0,3851	0,0960	0,1854

The bimodal behavior observed in t_o (see Figure 6.c) deserves further attention regarding gender issues. The glottal source modeling has been usually based on the L-F pattern, which is a good representation for the phonation cycle in male voice, but this may not be the case in female voice where vocal fold tension is larger, especially in young women [12]. It is well known that the contact phase is shorter in female voice, due to a larger tension in the folds and a lower dynamic mass, which produces a faster separation of the folds during the open phase start. Hormonal effects in larynx structures influence the developing of smaller vocal folds and tighter arytenoid stretching and cover tension. Hormonal decay reduces these effects. Cyclical and circumstantial changes in hormonal activity, as well as the effect of certain medication drugs may also alter the biomechanical status of the vocal folds. These observations indicate that the glottal source pattern in women deserves further attention and study.

V. CONCLUSIONS

The work presented give some answers to the questions posed in the introduction on the capability of wavelet patterns to describe the DCG, and opens new questions regarding the applications of the proposed methodology. These are some of the conclusions derived in this respect:

- The over-simplified vocal-fold/tract model seems to be adequate for establishing relations among dynamic variables and glottal gap. Of course, the vocal tract model may be described more accurately to include variable section tubes, source-filter interaction, and nasal effects, extending it to a more complex description derived from [8], although this first approximation gives a view of the glottal source behavior in deficient closure conditions.
- The wavelet-like behavior of gap patterns in the glottal source may be justified on these bases. From the distributions for male and female normophonic speakers the detection accuracy of the true and false opening instants and the DPG seems to be acceptable, as these agree with overall phonation features observed in both

genders (DPG is known to be larger in women than in men, contact phase is known to be shorter in women than in men, etc.).

- An important result is the capability of t_{og} , t_o and DPG for gender detection, as it seems that these parameters distribute differently for both genders (see Figure 6.a, b and d), with little overlap in their respective histograms. Although this circumstance has to be assessed by distribution hypothesis tests, it seems that these parameters may be used in gender detection. An important derived line of future research also in this sense is to carry out a study on feature distributions accordingly with age as well [13], as aging-related features may play a relevant role in neurologic disease detection and prevention.
- Another important study left pending is the possibility of opening a research line on female phonation and hormonal changes. Apparently the bimodal distribution of t_o may be explained assuming that the female population randomly selected for the normophonic female database may reflect the influence of hormonal activity in phonation [14], having in mind that the female subjects examined were in the range from 20 to 45 years. Thus almost half the population was showing a tensor voice (28 cases out of 50), as expected, and therefore a short contact phase (first mode in Figure 6.b) whereas 22 cases were showing a larger contact phase, more similarly to the male distribution. It is well known [16][17][18] that hormonal changes may induce liquid retention in vocal folds, increasing the dynamic mass, and possibly reducing tension, which would explain the extension of the contact phase in percent relative to the total glottal cycle. This hypothesis is being studied thoroughly and new results are expected in the future.
- From the results presented in TABLE II it seems also reasonable to hypothesize that t_o may be a good mark for pathologic voice detection and grading, as this parameter is well below the Q_1 - Q_3 range in male as well as in female voice in most of the pathological cases reported in table TABLE II. A reduction in the opening time may be explained by vocal fold pathology-induced over-tension, which could be a correlate to many diseases altering the behavior of the vocal fold cover.

As a final remark it must be said that the accurate estimation of the DPG and the DCG is an important objective in itself to progress towards automatic larynx pathology classification from acoustic measurements. Basically the patterns of pathology behavior in vocal fold vibration can be distributed between contact closure defects (contact extension, as edemae, or contact reduction, as nodules, polyps, etc) and asymmetric behavior (unilateral cysts, polyps, etc.). The marks provided by contact defects may be good descriptors pointing to specific pathologies. Besides, these may be also very important biometric marks, as these features have a permanent effect on phonation, and are less dependent on phonation modalities than other biomechanical parameters. This is also a very important research line to be pursued in the near future. Of course, the main conclusion is that the study of gap defect modeling and characterization is a very open field, needing more generalization. A full validation study including a large database of pathological voice may serve to validate some of

the preliminary conclusions exposed here, which have to be taken as exploratory postulates needing statistical validation and contrast. Another research task left for future studies is the validity of different wavelets in characterizing better contact defects over the extension presented. Practical applications of this methodology are the design of diagnose-aids for laryngologists and speech therapists, as well as for the educators of the singing voice. Pre- and post-treatment tests based on this methodology may grant a full monitoring method of good practice and a guide to e-health systems in care of human phonation.

ACKNOWLEDGMENT

This work has been funded by grants TEC2009-14123-C04-03 and TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Economic Affairs and Competitiveness of Spain. The authors wish to thank the collaboration of Drs. Mario Fernández and Carlos Ramírez in the study of particular cases and in providing essential information on the databases collected at Hospital del Henares at Coslada, Spain.

REFERENCES

- [1] Bielamowicz, S., Kapoor, R., Schwartz, J., Stager, S. V., "Relationship among Glottal Area, Static Supraglottal Compression, and Laryngeal Function Studies in Unilateral Vocal Fold Paresis and Paralysis", *J. Voice*, Vol. 18, 2004, pp. 138-145.
- [2] Uloza, V., Kašėta, M., Pribušienė, R., Šaferis, V., Jokūžis, V., Gelžinis, A., Bačauskienė, M., "Quantitative microalalyngoscopic measurements of vocal fold polyps, glottal gap and their relation to vocal function" *Medicina (Kaunas)*, Vol. 44, No. 4, 2008, pp. 266-272.
- [3] Tao, C., Jiang, J. J., "The phonation critical condition in rectangular glottis with wide prephonatory gaps", *J. Acoust. Soc. Amer.*, Vol. 123, No. 3, March 2008, pp. 1637-1641.
- [4] Alku, P., "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering", *Speech Comm.*, Vol. 11, 1992, pp. 109-118.
- [5] Dubuisson, T., Drugman, T., Dutoit, T., "On the mutual information of glottal source estimation techniques for the automatic detection of speech pathologies", *Proc. MAVEBA 2009*, Firenze Univesity Press, Florence, Italy, 2009, pp. 53-56.
- [6] Behroozmand, R., Almasganj, F., "Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis", *Comp. in Biol. and Med.*, Vol. 37, 2007, pp. 474-485.
- [7] deVries, M. P., Schutte, H. K., Veldman, A. E. P., Verkerke, G. J., "Glottal flow through a two-mass model: Comparison of Navier-Stokes solutions with simplified models", *J. Acoust. Soc. Amer.*, Vol. 111, 2002, pp. 1847-1853.
- [8] Klatt, D. H. and Klatt, L. C., "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *J. Acoust. Soc. Amer.*, Vol. 87, 1990, pp. 820-857.
- [9] Fant, G., Liljencrants, J., Lin, Q., "A four-parameter model of glottal flow", *STL-QPSR*, 4/1985, pp. 1-13.
- [10] Mallat, S., *A wavelet tour of signal processing*, Academic Press, San Diego CA, 1998.
- [11] Gómez, P., Fernández, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira, L. M., Martínez, R., Godino, J. I., "Glottal Source Biometrical Signature for Voice Pathology Detection", *Speech Comm.*, Vol. 51, 2009, pp. 759-781.
- [12] Hirano, M. et al. "Acoustic analysis of pathological voice. Some results of clinical application.", *Acta Otolaryngologica*, Vol. 105 (5-6), 1988, pp. 432-438.
- [13] Hanson, M., "Glottal characteristics of female speakers: Acoustic Correlates", *J. Acoust. Soc. Amer.*, Vol. 101, 1997, pp. 466-481.
- [14] C. Muñoz, R. Martínez, P. Gómez, E. W. Lang, A. Álvarez, L. M. Mazaira and V. Nieto, "KPCA vs PCA Study for an Age Classification of Speakers, *Lecture Notes on Artif. Intell.*, Vol. 7015, 2011, pp. 190-198.
- [15] M. B. Wadnerkar, P. E. Cowell and S. P. Whiteside, "Speech across the menstrual cycle: A replication and an extension study", *Neurosc. Letters*, Vol. 408, 2006, pp. 21-24.
- [16] J. Baker, "A Report on Alterations to the Speaking and Singing Voices of Four Women Following Hormonal Therapy With Virilizing Agents", *J. Voice*, Vol. 13, No. 4, 1999, pp. 496-507.
- [17] Abitbol, J., Abitbol, P., and Abitbol, B., "Sex Hormones and the Female Voice", *J. Voice*, Vol. 13, 1999, pp. 424-446.
- [18] Wadnerkar, M. B., Cowell, P. E., and Whiteside, S. P., "Speech across the menstrual cycle: A replication and extension study", *Neuroscience Letters*, 408, 2006, pp. 21-24.