Evaluation of Voice Pathology Based on the Estimation of Vocal Fold Biomechanical Parameters


**†Madrid, Spain

Summary: Voice disorders are a source of increasing concern as normal voice quality is a social demand for at least one third of the population in developed countries in cases where voice is an essential resource in professional exercise. In addition, the growing exposure to certain pathogenic factors such as smoking, alcohol abuse, air pollution, and acoustic contamination, and other problems such as gastro-esophageal reflux or allergy as well as aging, aggravate voice disorders. Voice pathologies justify the assignment of larger resources to prevention policies, early detection, and less aggressive treatments. Traditional pathology detection relies on perceptive evaluation methods (GRABS), acoustic analysis, and visual inspection (indirect laryngoscopy, and modern fibro-endo-stroboscopy). This article describes a method for voice pathology detection based on the noninvasive estimation of vocal cord biomechanical parameters derived from voice using specific signal processing methods. Preliminary results using records from patients showing four frequent causes of voice pathology (nodules, polyps, chronic laryngitis, and Reinke’s edema) are given. The results show that the alteration (distortion, unbalance, or deviation) of cord biomechanical parameters may serve as an indicator of pathology. Statistical methods based on hierarchical clustering and principal component analysis reveal that combining biomechanical estimates with classic perturbation parameters increases the accuracy of acoustic analysis, improving the detection of voice pathology. This research could open new possibilities for noninvasive screening of vocal fold pathologies and could be used in the implantation of e-health voice care services.

Key Words: Speech pathology—Vocal fold biomechanics—Speaker biome-try—Glottal source—Mucosal wave.
INTRODUCTION

The field for voice pathology research has experienced a great push forward during the last years due to the development of better models of the voicing apparatus, the application of more advanced exploration devices, and the introduction of powerful signal processing and pattern recognition tools. The research on the field has been concentrated in the estimation and classification of voice parameters that can be used in the determination of the presence of pathology and the degree of affection to the normal voicing process. These methodologies have introduced important advances in the detection and treatment of voice pathologies and in improving voice care. The objective evaluation of the vocal function may be carried out through different observational and objective evaluation methods, including laryngoscopic, aerodynamic, glottographic, and acoustic analysis. Most of these methods have been widely employed in isolation or in a combined practice to test their reliability in differentiating nondysphonic from dysphonic voices. Nevertheless, the features observable by laryngoscopic analysis commonly associated with physiological dysphonia (PD) are frequently prevalent in the nondysphonic population and fail in helping to distinguish patients with PD from nondysphonic subjects. Auditory-perceptual evaluation of voice has been a usual touchstone for other objective methods, but besides not being free of subjectivity, it requires the actual presence of a trained listener to evaluate voice quality in a reliable way. This requirement is expensive, time consuming, and out of the scope of a screening program. It is clear that there is a never-ending controversy in the field of voice research among supporters of perceptual versus instrumental voice assessment methodologies, in which both parts may show important arguments supporting their respective points of view. Recently, a combination of acoustic and auditory-perceptual methods has yielded almost a 100% improvement in the accuracy of each method applied in isolation in the task of discriminating nondysphonic from dysphonic voices. It means that the wise use of both methodologies in cooperation may render significant progress in the detection of pathology. Acoustic measurements of the vocal function are routinely used in the assessment of disordered voice, and in monitoring the patient’s progress over the course of voice therapy. Despite the objective nature of acoustic measures, when performed over increasing dysphonic voices (less periodic signals), the results are characterized by increasing measurement errors. Such behavior seems to be inherent to commercial applications as it has been adequately documented. In addition, this kind of acoustic analysis has demonstrated poor performance, moderate reliability, and side effects such as increased measurement sensitivity to changing environmental conditions. Therefore, caution should be observed in the non-contrasted use of computer-based acoustic analysis as the only means to estimate voice quality in isolation from other methods. It is especially so in relation with clinical interventions aimed at improving voice quality. Recently, new parameters have been introduced and clinically tested for estimation of the turbulent noise in voice signals (turbulent noise index (TNI)) and for the “breathy” voice characterization (normalized first harmonic energy (NFHE)) for laryngeal pathology detection, with an identification accuracy of 96.1% using the K-nearest neighbor clustering methods. Neural-network-based classification approaches applied to the automatic detection of voice disorders like learning vector quantization yielded a 96% of accuracy detecting voice pathology. Aerodynamic measures have been frequently used to analyze pathologic voices, but accuracy in determining a voice to be dysphonic or nondysphonic has been reported to be 91.1% with a certain overlapping between nondysphonic and dysphonic groups.

Having in mind the limitations of previous research and other objective methods for voice evaluation, a new methodology for the detection of pathology has been proposed. This technique uses a combination of classic distortion and biomechanical parameters estimated from glottal dynamics correlates. From the analysis of the results, it is shown that this technique would be able to differentiate between nondysphonic and dysphonic voices and hopefully between different pathologies in the long term. A biomechanical approach to the
problem of pathology detection has the advantage over other methods of descending to the closer vicinity where physiology and physics of voice production engage (vocal fold: source). The approach intends to aim at the origins and foundations of the problem, removing supraglottal influences by searching for the relations between different types of vocal fold pathologies and their underlying pathophysiological mechanisms.

The purpose of the current research is to be established in the framework of the estimation and classification techniques of parameters obtained from the voice signal, concentrating specifically on those derived from the vocal fold dynamical theory, to:

1. Determine the suitability of vocal fold biomechanics in complementing classic distortion parameters in pathology detection.
2. Explore its capability for pathology determination.
3. Infer the possibility of offering the voice therapist specific hints to the biomechanical conditions of the patient’s vocal folds, pointing to help in pathology determination and classification.

Because these objectives are ambitious, the current paper will be devoted to describe in depth the research performed in connection with objective 0, leaving objectives 0 and 0 for further research. The specific questions to be treated and answered are the following:

1. Is it possible to define a certain signal correlate to the dynamic behavior of the vocal fold (cord body and cover), with emphasis on the mucosal wave (traveling wave on the lamina propria)?
2. How can we better define and estimate the cord body and cover-dynamics correlates?
3. How would vocal fold dynamics be reflected on the respective power spectral densities of both correlates?
4. How can we parameterize the power spectral density of these signals to optimally describe vocal fold dynamics?
5. Could the biomechanical parameters from the vocal fold dynamics be determined using spectral estimations of the glottal source correlate?
6. Could these biomechanical parameters or others derived from them complement classic distortion parameters in the detection of voice pathologies?

The structure of the paper is as follows: The definition of cord body and cover correlates and especially the mucosal wave as a traveling wave will be briefly reviewed from contributions by the researchers in the field. The three-mass model of the vocal fold body–cover structure will be introduced as a means to explain the behavior of the vocal cords in terms of explaining the behavior of the body-cover dynamics. This study will allow us to conclude that a one-mass model of the cord body can explain the average movement of the vocal folds when the mucosal wave is not taken into account. Based on the behavior of the one-mass model, a precise definition of the body-dynamics correlate can be established. The extraction of the glottal source correlate from the elimination of the vocal tract influence from voice by inverse filtering will be reviewed next, commenting on the implementation details used within the framework of the current study. The next step will be to define the mucosal wave correlate as the residual found when the body-dynamics correlate is removed from the glottal source correlate. The correctness of the definition will be proven on the evidence extracted from the spectral behavior of the mucosal wave correlate as explained using a two-mass model of the cover. The process to estimate the biomechanical parameters of the cord body will be reviewed next. The exposition of the fundamentals of the study will end with the definition of parameter unbalance and deviations to be used as distortion parameters in pathology detection. The materials and methods used in the statistical analysis of the parameter estimations obtained will be given, concentrating on the production of the voice database used in the study. The results obtained with this database will be analyzed using multivariate techniques. These techniques show that a distinction among pathologic and nondysphonic samples may be obtained using classic distortion parameters in cooperation with biomechanical parameter deviations. A discussion on the results produced will be followed by one on the conclusions derived from the study.
FUNDAMENTALS OF THE METHODOLOGY PROPOSED

In this section, the background of the methodology proposed in the study for the processing of voice signals is to be reviewed. This methodology is intended to obtain estimations of the glottal source correlate and its power spectral density profile and singularities. From these estimations, basic biomechanical parameters are to be deduced. Classification techniques will be used to establish the capabilities of these parameters in determining the presence of pathology in voice.

The mucosal wave as a traveling wave

The first question deals with the definition of the mucosal wave correlate. Classically, it is well established that the movement of the vocal folds (Figure 1A) may be observed as the superposition of two main components: the dynamics of the vocalis muscle bulk tissue (body) and the one of the epithelium and superficial lamina propria (cover), the latter specifically known as the mucosal wave. This process is described as a traveling wave propagating over the cover tissue during the phonation cycle, consisting of a displacement of the tissues relative to the body. This displacement behaves as a cyclic perturbation that affects mostly the supraglottal and subglottal lips of the vocal cord. The closure phase would be produced by full contact between both vocal cords resulting from perturbations produced by the traveling waves on the respective cover structures.11–13

The three-mass model of the vocal fold body-cover structure

The behavior of the mucosal wave as a traveling wave phenomenon may be explained according to the dynamics of the body and the three-layer structure of the cover by means of biomechanical models of the vocal fold system. There have been numerous attempts to develop mechanical models reproducing the physics of phonation in the laboratory and through computer simulations. These attempts go from the classic two-mass model of Ishizaka and Flanagan to the more complex computational models based on finite differences.17–21 Many efforts have been made to employ vocal cord models in explaining the detailed vibration of the vocal fold system to extract knowledge applicable onto different fields, such as voice production, processing, and clinical studies. The use of k-mass vocal fold models helps in determining and characterizing the main components of vocal fold movement, especially the mucosal wave effect.13,19 For the purposes of this research, an equivalent k-mass model (Figure 1B) has been used. Here body dynamics is represented by a large lumped mass positioned at the equivalent center of masses of the body tissue structure, and the cover is represented by a set of k – 1 lumped masses linked by springs among themselves and to the body mass. Each parameter should be considered a representation of a specific dynamic situation describing the whole vibration process at a given time instant. This representation implies that a specific distribution of masses and tensions will give a description

FIGURE 1. A. Cross-section of the left vocal cord showing the body and cover structures (taken from Ref. 10). B. k-mass model of the body and cover, this last structure being modeled as a set of masses linked by springs among themselves and to the body mass.
of a certain dynamical state of the system and may change accordingly (time-varying model). Each spring will be represented by a stiffness parameter, and a loss factor will be associated with every mass to account for viscous and other losses. The complexity of this model will be adapted for the purposes of the research introducing some classic simplifying assumptions:

1. Each vocal cord, a three-dimensional (3-D) complex structure, will be considered if it is vibrating uniformly along its longitudinal dimension (orthogonal to the section shown in Figure 1A, on the z axis orthogonal to the plane defined by the axes x and y).

2. Any lumped mass in the model would be the equivalent dynamical representation of the forces and accelerations acting on a portion of the equivalent 3-D structure such as that concentrated on its ideal center of masses.

3. Model masses would move only along the axis x, as a result of forces applied solely in the same axis.

4. The changes in the biomechanical parameters with time should be slow enough compared with the time interval of study, assuming that the model is a time invariant system.

When the total number of masses is reduced to 3, the resulting model is the well known Story–Titze’s three-mass model, which is the simplest structure that can approach the representation of the mucosal wave effect “maintaining the simplicity of a low-dimensional system.” This model, which is represented in Figure 2, gives a good description of the body-cover dynamics and as such has been widely studied. The body masses \( M_{bl} \) and \( M_{br} \) are linked to the reference rigid walls of the thyroid and arytenoid cartilages by equivalent springs with elastic parameters \( K_{bl} \) and \( K_{br} \). The cover structure is represented by two masses per cord (\( M_{il} \) and \( M_{jl} \) for the left cord and \( M_{ir} \) and \( M_{jr} \) for the right one) that are linked to the respective body masses by springs \( K_{il}, K_{jl}, K_{ir}, \) and \( K_{jr} \) and among themselves by \( K_{ijl} \) and \( K_{ijr} \). The upper and lower masses would represent the differentiated movement of the subglottal and supraglottal tissues of the cover. The dynamic variables are the forces acting on the respective cover masses \( f_{xij} \) (\( i, j \) designating the variables associated with the subglottal and supraglottal regions of the vocal fold) as a result of the supraglottal to subglottal pressure difference, and the velocities associated with each mass, \( v_{bl,r} \), \( v_{il,r} \), and \( v_{jl,r} \) (where \( b \) designates the parameters associated with the cord body, and \( l \) and \( r \) refer to the

FIGURE 2. Schematic structure of one section of the cover modeled as a three-mass system. This is a version of the well-known Story–Titze model, which is agreed to give a more accurate explanation of the mucosal wave phenomenon than Ishizaka–Flanagan’s two-mass model. Under the adequate assumptions, this model may be reduced to a one-mass model (through its equivalent in Figure 11), taking into account average body dynamics only, or the two-mass equivalent model of the cover dynamics (when referenced to the body masses). These two reductions have been used in the study to derive estimates of the body biomechanical parameters and to compare the spectral behavior of the mucosal wave correlate with the two-mass model response as given in Figure 10.
left and right vocal folds). The differentiation between left and right vocal folds may seem superfluous when both vocal folds are considered symmetric, but this assumption does not stand even for normal voicing, as morphologic and dynamic differences between both cords may be appreciated even in nondysphonic speakers. A first approach to the study of the vocal cord dynamics would consider it described by the movements of the two cover masses $M_{il,r}, M_{jl,r}$ on each cord relative to the respective body masses $M_{bl,r}$. The dynamical behavior of the vocal cords would be represented by a signal that can be referred to as the glottal pulse (flow) correlate reconstructed after removing the vocal tract influence from voice.\textsuperscript{23,24} The first derivative of the glottal correlate can be observed on its turn as the glottal source correlate and is indirectly related to the relative speed between the cord centers of mass.\textsuperscript{25–27} The influence of the body and cover dynamics on the glottal source correlate would appear as two different contributions: a slow-varying average movement, also known as the average acoustic waveform\textsuperscript{3} or average glottal source correlate, and a fast-varying waveform, presumably induced by the mucosal wave traveling on the body-cover structure,\textsuperscript{11,12} which will be referred to as the cover-dynamics correlate.

**The one-mass model of the cord body**

As a result of the current research, a separated study of body and cover dynamics is assumed. For such, the effects of the body dynamics and the cover dynamics are to be independently estimated from the glottal source correlate. It should be expected that the power spectral density of the average glottal source correlate would be determined by the dynamics of the cord center of masses (body dynamics). On its turn, the power spectral density of the cover-dynamics correlate would be mostly influenced by the behavior of the cover-dynamics structures. Under such an assumption, the average glottal source correlate would coincide with the vibration pattern of a pair of vocal cords modeled as a one-mass system. This system may be considered a simplification of the Story–Titze model, considering that stiffness parameters $K_{il,r}, K_{jl,r}$ and $K_{ijl,r}$ are indefinitely large; thus, the set of three-masses per cord would reduce to a single point-like global lumped mass $M_{gl,r}$ given by

$$M_{gl,r} = M_{bl,r} + M_{il,r} + M_{jl,r} \quad (1)$$

This single mass is tied to the rigid walls of the glottis by the springs $K_{gl,r}$, the dynamic equation of the reduced 1-mass structure being given as

$$f_{cl,r} = M_{gl,r} \frac{\partial^2 v_{gl,r}}{\partial t^2} + R_{gl,r} v_{gl,r} + K_{gl,r} \int_{-\infty}^{t} v_{gl,r} dt \quad (2)$$

The equivalent global masses $M_{gl,r}$ of the vocal folds vibrate as a response to the force $f_{cl,r}$ exerted by the supraglottal and subglottal pressure differences. The vibration along the axis $x$ will be described by the position of the global masses $M_{gl,r}$ at each instant (Figure 3A). This movement when plotted versus time would be described by arch-like oscillation cycles, which are a result of ideally elastic cord collision, represented by the rectified sinusoids with frequencies $\omega_{pl}$ and $\omega_{pr}$ given by the resonant system composed by both cords as

$$\omega_{pl} = \sqrt{\frac{K_{bl}}{M_{gl}}} \quad \omega_{pr} = \sqrt{\frac{K_{br}}{M_{gr}}} \quad (3)$$

The resulting vibration patterns would be given by plots similar to the ones shown in Figure 3, these being the result of considering the cords as vibrating freely, colliding once in a vocal cycle, and separating again in a bouncing movement. The behavior of the cords between collisions would be given by the dynamics of the one-mass model as given in Equation 2.

**Estimating the body and cover-dynamics correlates from the glottal source**

The reconstruction of the body and cover-dynamics correlates is based on the glottal source correlate obtained by inverse filtering of voice.\textsuperscript{23,24,28–30} Figure 4 shows the specific use of this methodology in the reconstruction of the glottal source correlate within the current study. In step 1, the input voice $s(n)$ is filtered using an inverse radiation model filter $H_f(z)$ to compensate the radiation effects at the lips to produce a trace of radiation compensated voice $s_f(n)$. In step 2, a simple glottal pulse inverse model

\[ \text{Journal of Voice, Vol. 21, No. 4, 2007} \]
filter $H_g(z)$ is used to cancel the behavior of the glottal source in the radiation compensated voice, producing a trace of de-glottalized voice $s_v(n)$. This signal is inverse-filtered in step 3 using lattice filters to extract the model of the vocal tract given by the transfer function $F_v(z)$. The inverse of this function is applied in step 4 to the radiation compensated voice $s_l(n)$, producing a residual trace containing only information on the glottal source derivative $v_g(n)$. In step 5, this signal is modeled also by inverse filtering to reconstruct its equivalent generation model $F_g(z)$. Its inverse function $H_g(z)$ is used to

![Image](image-url)  
**FIGURE 3.** Vibration patterns of vocal cords represented by the one-mass model. **A.** Position of the right cord body mass (full line), and position of the left cord body mass (dot line). **B.** Right cord mass velocity. **C.** Speed of the body mass relative to its resting position ($x = 0$). The vibration cycle is as follows: (1) Both left and right masses start separating from the resting position. (2) Maximum separation is reached, the relative speed becomes null, the cord tension inverts the movement, the velocity becomes negative in the right cord mass, and positive in the left cord mass. (3) Both masses come in close contact (possible collision effects) during a small fraction of time to separate again and start a new vibration cycle (4).

![Diagram](diagram-url)  
**FIGURE 4.** Iterative implementation of Alku’s method used in the current study. (1) Elimination of the lip-radiation effects. (2) Elimination of the glottal source spectral fingerprint on input voice. (3) Estimation of the vocal tract transfer function by inverse linear predictive filtering using adaptive paired lattices. (4) Elimination of the vocal tract transfer function on input voice. (5) Estimation of the glottal source transfer function to be applied in 2. The process is iterated 2-3 times to obtain a refined residue $v_g(n)$, which is shown to be a correlate of the glottal source first derivative.
refine the estimation of $s_i(n)$ by repeating step 2 and, consequently, steps 3–5 in a loop iteration as many times as necessary (typically two to three are enough to improve the accuracy of inverse filtering to a reasonable limit). Through successive iterations, glottal source derivative $v_g(n)$ (a refined estimation of the second derivative of the glottal source) and the deglottalized voice trace $s_v(n)$ (impulse response of the vocal tract) are obtained concurrently. This last trace is of special relevance in the robust estimation of the formants (resonances) of the vocal tract after removing the influence of the glottal dynamics as shown in Figure 5. The templates shown in Figure 6 give a view of the results of the inverse filtering when applied to a specific voice trace (Figure 6A). In Figure 6B, the correlate of the second derivative of the glottal source $v_g(n)$ is obtained (compare its resemblance with the trace in Figure 3C corresponding to sample case “00B”). The integration of $v_g(n)$ results in the trace shown in Figure 6C, which should be associated with the glottal source correlate. Finally Figure 6D shows the results of a new integration to reproduce the correlate of the glottal pulse $u_g(n)$. In Figure 7A and C, both the glottal source and the glottal pulse have been referenced to a common ground level. It is well known that linear prediction-based inverse filtering represents a system as an all-pole filter by constructing a FIR filter implementing the inverse transfer function of the system. Therefore, systems that do not behave as all-pole filters may be deficiently modeled using linear prediction. Based on this fact, the methods used in this study for inverse filtering followed a careful design:

1. The structure of the filters used is a variant of linear predictive filters known as lattice filters, which reproduce the tube structure of the vocal tract by means of associated reflection coefficients. An adequate sampling rate produces a precise slicing of the tube structure into smaller subsections, producing accurate estimations of the forward and backward waves propagating in the vocal tract. These estimations allow a more precise reconstruction of the pressure and flow variables along

![FIGURE 5. LPC (all-pole) power spectral densities of the signals in the inverse filtering process shown in Figure 4: A. The input voice trace $s(n)$ showing the formant structure of voice. B. The glottal source derivative $v_g(n)$ showing a smooth decay of about –6 dB/oct. C. The de-glottalized voice $s_v(n)$ conveying the formant structure due to the tube structure of the vocal tract.]

2. Adaptive versions of lattice filters have been used; therefore, the inherent noninvariance of the dynamic behavior of vocal tract and vocal fold vibration has been carefully traced.

3. Using the properties of paired lattices, the concurrent determination of the all-pole structure of the system modeled has been combined with the removal of its inverse transfer function from the respective cross-signals in the flow graph shown in Figure 4.

4. The iterative estimation of the glottal and vocal transfer functions allows a progressive refinement of the estimates, which may be shown empirically to converge to a stable solution after a few iterations in all cases studied.

5. Special care has been placed in granting that inverse filtering was applied only on voiced sounds associated with all-pole vocal tract transfer functions, as non-nasalized vowels.

The separation of the average glottal source and the cover-dynamics correlate from the glottal source correlate could be carried out in different ways, such as low-pass filtering combined with wave rectification. Several methods have been used in different preliminary studies, such as low-pass filtering or signal unfolding and subtraction. The results presented here correspond to the subtraction of...
sinusoidal arches with the same semi-period as the glottal aperture, on a phonation cycle basis. An adaptive method to evaluate the amplitude of the sinusoidal arch is used, based on the minimization of the energy of the error between the glottal aperture \( y_g(n) \) and the sinusoidal average \( s_g(n) \) as

\[
L = \sum_{n \in N_k} e^2(n) = \sum_{n \in N_k} (y_g(n) - s_g(n))^2 \tag{4}
\]

where \( N_k \) is the set of samples associated with the \( k \)-th cycle of phonation, and

\[
s_g(n) = y_{0k} \sin(\omega_k n \tau) \quad n \in N_k \tag{5}
\]

\( y_{0k} \) being the amplitude of the sinusoidal arch to be estimated minimizing the cost function \( L \) as

\[
\frac{\partial L}{\partial y_{0k}} = 0 \quad \Rightarrow \quad y_{0k} = \frac{\sum_{n \in N_k} y_g(n) \sin(\omega_k n \tau)}{\sum_{n \in N_k} \sin^2(\omega_k n \tau)} \tag{6}
\]

The cover-dynamics component or mucosal wave correlate is defined as the difference between the glottal aperture and its approximation by semi-sinusoidal arches as

\[
y_{ck}(n) = y_g(n) - y_{0k} \sin(\omega_k n \tau) \tag{7}
\]

From this last expression, it may be inferred that the mucosal wave correlate has been defined as the minimum energy signal obtained by subtracting the body-dynamics component from the glottal source correlate. Figure 7B and D shows the body and
cover-dynamics components, respectively. In Figure 8, the results of estimating the cover-dynamics component during a phonation cycle are shown in a closer view. It may be appreciated that in this case, a large amount of the higher vibration modes is contributed by the third harmonic. The respective power spectral densities of the glottal source, its average component, and the mucosal wave correlate are given as well in Figure 9.

The two-mass model of the cover

The consistency of the mucosal wave correlate definition may be established by comparing the spectral behavior of its power spectral density against the trans-admittance of the two-mass cover model. This model may be derived from the three-mass model of Figure 2, which refers all dynamic variables of the cover masses relative to the body mass. This assumption would allow us to separate the dynamics of the cover from that of the body. The cover structure could then be modeled with the two-mass Ishizaka–Flanagan model. In Figure 10A, the power spectral density of the mucosal wave correlate extracted from a nondysphonic subject has been plotted against frequency. As a contrast, the square modulus of the inter-mass trans-admittance of the equivalent two-mass cover model has been plotted in Figure 10B. (This trans-admittance is the equivalent transfer function between the speed measured in one mass of the cover when the other mass is excited with a sinusoidal force, arresting any excitation at the first mass. It gives a measurement on how dynamic excitation propagates from mass to mass on the cover.) It may be observed that adjusting adequately the values of the biomechanical parameters of the two-mass model ($M_i, M_j, K_i, K_j$, and $K_{ij}$, assuming the vocal cords to be symmetric in what follows), the trans-admittance reproduces the main characteristics of the power spectral density of the mucosal wave correlate: sharp rise from low frequencies to a first maximum ($T_{m1}, f_{m1}$), deep decrease to a trough ($T_m, f_m$), and a second peak defined by a second resonance ($T_{m2}, f_{m2}$). This v-grooved structure is due to the interaction between the two masses and the dynamic properties of the cover.
to the interaction of the cover masses linked by springs like $K_{ij}$. The decay toward high frequencies shows the presence of other less relevant mass-spring substructures marked by other v-grooved troughs. The behavior of such a structure is theoretically described in Berry\textsuperscript{22} and well documented experimentally in the work by Švec et al.\textsuperscript{32} In this last work, the authors measure the spectral density of the precise movement of a specific point in the supraglottal ridge of both vocal cords, revealing the v-grooved pattern being discussed, which shows a clear relation with the cover dynamics modeled by multiple masses linked by springs.

Estimating the biomechanical parameters of the cord body

In this study, emphasis is placed on the estimation of the body biomechanical parameters, leaving an equivalent study on the cover parameters for future work. This estimation is related to the inversion of the integro-differential equations of the one-mass vocal fold model as given in Equation 2, by synthesizing the elements of its equivalent electromechanical circuit in the frequency domain as given in Figure 11. For such, estimates of the magnitudes involved in body dynamics—essentially from the envelope of the glottal source power spectral distribution—are to be used. Conceptually the problem may seem simple, the underlying difficulties being found in the definition of robust estimates of the mentioned magnitudes.

It has been shown in a previous work\textsuperscript{33} that reliable estimates of the relative values of cord body masses and tensions could be obtained from the power spectral density of the average acoustic waveform. The estimation technique used was the adaptive fitting of the power spectral density against the transfer function of the one-mass model. The work hypothesis\textsuperscript{26,27,33} is based on the assumption that the envelope of the power spectral density

![Figure 9](https://example.com/figure9.png)

**FIGURE 9.** Power spectral densities of (A) the glottal source correlate, (B) the average glottal source, and (C) the cover-dynamics component or mucosal wave correlate. The vertical axes are given in decibels for all templates.
of the body-dynamics component is directly related with the square modulus of the input admittance derived from Equation 2 as

\[ T_b(\omega) = \frac{|Y_b|^2}{|F_x|^2} = \left( \frac{\omega M_b - \omega^{-1} K_b}{\omega M_b - \omega^{-1} K_b + R_b} \right)^2 \]

where \( M_b \), \( K_b \), and \( R_b \) are, respectively, the parameters associated with the lumped mass, elasticity, and losses of the one-mass model when only the body of the vocal cord is taken into account following the reduction of Story–Titze’s model as explained in previous sections. \( F_x(\omega) \) and \( V_x(\omega) \) are the frequency domain estimations of the dynamic variables of the cord body (force and speed along the \( x \) axis, respectively).

The robust estimation of the model parameters is based on the precise determination of two points on the power spectral density of the cover-dynamics component, these being \( \{ T_{b1}, \omega_1 \} \) and \( \{ T_{b2}, \omega_2 \} \). The lumped body mass may be estimated as

\[ M_b = \frac{\omega_2}{\omega_2^2 - \omega_1^2} \sqrt{\frac{T_{b1} - T_{b2}}{T_{b1} T_{b2}}} \]

The selection of the most adequate points for \( \{ T_{b1}, \omega_1 \} \) and \( \{ T_{b2}, \omega_2 \} \) is not a trivial issue, as it is highly related with the robustness of the estimation procedure. A good candidate for \( \{ T_{b1}, \omega_1 \} \) is the position of the main (resonant) peak in the amplitude of the power spectral density of the glottal source. The peak position is apparently easy to determine, but it is contaminated by the sampling capability of the short-time Fourier transform (STFT).

The use of STFT is compulsory if a dynamic study is to be carried out on a phonation cycle basis, as it

![Figure 10](image1.png)

**FIGURE 10.** A. Power spectral density of the mucosal wave correlate as defined in Equation 7 for nondysphonic voice (sample “035”). A trough of amplitude \( T_n \) at a frequency \( f_n \) is surrounded by two peaks \( t_{m1}, f_{m1} \) and \( t_{m2}, f_{m2} \). This v-grooved pattern is associated with the two cover masses and the spring linking them, as inferred from the study of Berry on the two-mass model. B. Square modulus of the trans-admittance relating the force acting on one of the cover masses and the linear speed induced in the other cover mass when no force is driving it. The system is supposed to be under sinusoidal permanent regime. The vertical axes are given in decibels for both templates.

![Figure 11](image2.png)

**FIGURE 11.** One-mass electromechanical equivalent of the body of one of the vocal cords (assumed symmetric). \( f_x \) and \( v_x \) are the respective force and speed on the body mass \( M_b \) along the \( x \) axis. \( R_b \) and \( K_b \) are the respective parameters of losses and stiffness.
is in the current case. Typically, if the pitch is around 200 Hz (such as in female voice), the phonation cycle may last 5 ms. Using a sampling frequency of 44,100 Hz, this means 220 samples, which implies a resolution in the frequency domain of 100 Hz. The determination of pitch position and amplitude will be affected by this low resolution. The length of the cycle window may oscillate typically between 400 (male voice) and 200 (female voice) samples per cycle for most cases. The low-resolution problem could be improved if larger sampling frequencies (above 100 kHz) were used. Interpolation by splines on the power spectral density obtained from STFT is a technique that works reasonably well in obtaining robust estimates of the peak position and amplitude. A good candidate for \( \{ T_k, \omega_k \} \) is the position of the third harmonic from the peak position, as the time series shows odd symmetry. These two points have shown to be robust enough for the cases under study, although other strategies are also possible. Once the mass has been estimated, the elastic parameter (body stiffness) \( K_b \) may be obtained from the precise determination of the position of the resonant peak (after applying splines), this being \( \{ T_r, \omega_r \} \)

\[
K_b = M_b \omega_r^2
\]

the parameter of body losses being estimated (but for a scale factor \( G_b \)) as

\[
R_b = \frac{G_b}{\sqrt{T_r}}
\]

where \( T_r \) stands for the value of the square modulus of the input admittance in Equation 8 at the frequency of resonance \( \omega_r \). Estimations of the cord body biomechanical parameters obtained for a phonation cycle from a nondysphonic voice trace (subject sample “00B”) have been used to reconstruct the square modulus of the admittance following Equation 8, which is presented against the power spectral density of the glottal source correlate in Figure 12 for comparison. The estimation of \( G_b \) is carried out normalizing the glottal source correlate with respect to the maximum amplitude of the average glottal source among all phonation cycles contained in the voice frame under analysis. The estimation of \( G_b \) is only relevant if the objective is to obtain absolute estimates of the biomechanical parameters. For the purpose of pathology detection, only relative estimates of the parameters are required as shown in the next paragraph; therefore, the role played by \( G_b \) is completely irrelevant in the current study.

**Estimating the parameter unbalance of the cord body**

Until now, it has been considered that for the purpose of spectral estimation and fitting, both vocal cords were symmetric. This assumption does not stand in most cases whether dysphonic or nondysphonic voice is involved. As a matter of fact, asymmetry will result in the unbalance of the biomechanical parameters estimated for neighbor phonation cycles. It seems reasonable to think that this unbalance will be larger in cases where vocal fold pathology is present than in nondysphonic cases. Whether unbalance in vocal fold vibration is caused by physiological, psychological, or dynamical reasons, it will leave a fingerprint on biomechanical parameter estimations from a given subject when compared with phonation cycle by cycle. Leaving the study of the specific influence of body biomechanical parameter asymmetry on phonation cycle unbalance as an open research line, it will nevertheless be concluded that the presence of unbalance is generally accepted to be a correlate to vocal fold pathology (as related in a certain way with shimmer). Unbalance between neighbor phonation cycles may be appreciated in Figure 7 where the waveforms of the dynamic variables plotted tend to exhibit slight differences, which are more notorious in the case of the glottal pulse. The estimations of mass, stiffness, and losses are produced on a phonation cycle-frame basis; therefore, the (intra-speaker) unbalance of these parameters (BMU, body mass unbalance; BLU, body losses unbalance; and BSU, body stiffness unbalance) may be defined as

\[
\mu_{uk} = \frac{\hat{M}_{bk} - \hat{M}_{bk-1}}{\hat{M}_{bk} + \hat{M}_{bk-1}}
\]

\[
\rho_{uk} = \frac{\hat{R}_{bk} - \hat{R}_{bk-1}}{\hat{R}_{bk} + \hat{R}_{bk-1}}
\]

\[
\gamma_{uk} = \frac{\hat{K}_{bk} - \hat{K}_{bk-1}}{\hat{K}_{bk} + \hat{K}_{bk-1}}
\]

where \( 1 \leq k \leq K \) is the phonation cycle window index and \( \hat{M}_{bk}, \hat{R}_{bk}, \) and \( \hat{K}_{bk} \) are the k-th cycle estimates of mass, losses, and stiffness on a given
voice sample (for a single specific subject, i.e., intraspeaker).

Another interesting comparison would be the contrast of specific average values for a given subject with respect to a representative population of subjects, taking into account the presence or absence of a given pathology, age, sex, or other specific conditions. The deviations of the average estimates of mass, losses, and stiffness, over the set of phonation cycles contained in the frame under study from a given subject, relative to average estimates from a nondysphonic set of speakers (interspeaker), may be a good indicator of pathology as well. If $M_{bj}$, $R_{bj}$, and $K_{bj}$ are the phonation-cycle average estimates for subject $j$-th, and $M_{bs}$, $R_{bs}$, and $K_{bs}$ the averages of the same estimates over a nondysphonic population set, the deviation estimates could be defined as

$$m_{dj} = \frac{(M_{bj} - M_{bs})}{M_{bs}}$$
$$\rho_{dj} = \frac{(R_{bj} - R_{bs})}{R_{bs}}$$
$$\gamma_{dj} = \frac{(K_{bj} - K_{bs})}{K_{bs}}$$

which will be referred to in the sequel as the body mass deviation (BMD), body losses deviation (BLD), and body stiffness deviation (BSD).

**MATERIALS AND METHODS**

**Recording protocol and data collection**

All voice samples analyzed in the study were obtained in the Laboratorio de Biomecánica del Aparato Fonador (LABAF) as part of a screening program for voice disorder detection in the population of Madrid (Spain). Common international bioethical protocols for clinical investigation have been considered within the study (European Agreement for Bioethics). Both nondysphonic and dysphonic voice donors accepted voluntarily to participate in the study. The recording protocol included a description of voice history, an auditory-perceptual evaluation (GRABS scale), a telelaryngoscope video-stroboscopic exploration, an acoustic analysis, an electro-glottographic evaluation, and a voice recording. A long read paragraph...
(equivalent to the *Rainbow Passage* for Spanish speakers), the phonetically balanced sentence *les hâbil un solo dial*, and three sustained realizations of the vowel /al/ were acquired from each subject and stored. Vowel traces were 3 s long, excluding onsets and trails.

Subjects were classified after evaluation and discussion as nondysphonic or dysphonic based on videostroboscopy evidence, acoustic analysis, and electro-glottographic trace inspection. For a speaker to be classified as nondysphonic, each of these three tests should be passed without any evidence of vocal disorder. A senior laryngologist and a senior speech pathologist independently evaluated the figures in the sequel, the number of phonation cycles produced was 36 of phonation cycles produced was 36. The first two parameters estimated are the pitch ($p_1$, Figure 13) and jitter ($p_2$, Figure 14). Parameters $p_{3-6}$ are three variants of shimmer estimated by different algorithms (Figure 15). Parameters $p_{7-9}$ are related with the drifts in the glottal source and the duration of the closure interval (Figure 14). Parameter $p_8$ is the ratio between the energy of the average glottal

<table>
<thead>
<tr>
<th>Trace Condition</th>
<th>BMD</th>
<th>BLD</th>
<th>BSD</th>
<th>BMU</th>
<th>BLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>001 N</td>
<td>−0.632</td>
<td>−0.136</td>
<td>−0.540</td>
<td>0.027</td>
<td>0.039</td>
</tr>
<tr>
<td>003 N</td>
<td>−0.154</td>
<td>−0.145</td>
<td>−0.137</td>
<td>0.079</td>
<td>0.056</td>
</tr>
<tr>
<td>005 N</td>
<td>−0.039</td>
<td>−0.299</td>
<td>−0.213</td>
<td>0.078</td>
<td>0.044</td>
</tr>
<tr>
<td>007 N</td>
<td>−0.492</td>
<td>−0.461</td>
<td>−0.573</td>
<td>0.036</td>
<td>0.046</td>
</tr>
<tr>
<td>00A N</td>
<td>−0.542</td>
<td>−0.207</td>
<td>−0.567</td>
<td>0.065</td>
<td>0.064</td>
</tr>
<tr>
<td>00B N</td>
<td>1.320</td>
<td>0.642</td>
<td>1.250</td>
<td>0.149</td>
<td>0.191</td>
</tr>
<tr>
<td>00E N</td>
<td>−0.054</td>
<td>0.012</td>
<td>−0.128</td>
<td>0.159</td>
<td>0.098</td>
</tr>
<tr>
<td>010 N</td>
<td>−0.408</td>
<td>0.164</td>
<td>−0.491</td>
<td>0.115</td>
<td>0.103</td>
</tr>
<tr>
<td>018 N</td>
<td>−0.031</td>
<td>−0.205</td>
<td>−0.167</td>
<td>0.078</td>
<td>0.076</td>
</tr>
<tr>
<td>01C N</td>
<td>−0.557</td>
<td>−0.315</td>
<td>−0.581</td>
<td>0.058</td>
<td>0.052</td>
</tr>
<tr>
<td>024 N</td>
<td>0.631</td>
<td>1.330</td>
<td>1.200</td>
<td>0.120</td>
<td>0.124</td>
</tr>
<tr>
<td>029 N</td>
<td>0.101</td>
<td>−0.111</td>
<td>0.416</td>
<td>0.057</td>
<td>0.048</td>
</tr>
<tr>
<td>02C N</td>
<td>−0.329</td>
<td>−0.253</td>
<td>−0.079</td>
<td>0.035</td>
<td>0.040</td>
</tr>
<tr>
<td>02D N</td>
<td>−0.227</td>
<td>−0.193</td>
<td>0.022</td>
<td>0.116</td>
<td>0.053</td>
</tr>
<tr>
<td>032 N</td>
<td>−0.507</td>
<td>−0.019</td>
<td>−0.367</td>
<td>0.038</td>
<td>0.071</td>
</tr>
<tr>
<td>035 N</td>
<td>0.424</td>
<td>−0.302</td>
<td>−0.021</td>
<td>0.099</td>
<td>0.065</td>
</tr>
<tr>
<td>043 N</td>
<td>0.219</td>
<td>0.156</td>
<td>0.466</td>
<td>0.059</td>
<td>0.030</td>
</tr>
<tr>
<td>047 N</td>
<td>−0.497</td>
<td>1.070</td>
<td>−0.180</td>
<td>0.076</td>
<td>0.052</td>
</tr>
<tr>
<td>049 N</td>
<td>−0.157</td>
<td>0.160</td>
<td>0.029</td>
<td>0.113</td>
<td>0.079</td>
</tr>
<tr>
<td>04A N</td>
<td>−0.005</td>
<td>1.770</td>
<td>0.073</td>
<td>0.098</td>
<td>0.075</td>
</tr>
<tr>
<td>065 BP</td>
<td>0.240</td>
<td>7.490</td>
<td>3.220</td>
<td>0.835</td>
<td>0.712</td>
</tr>
<tr>
<td>069 LVCP</td>
<td>0.560</td>
<td>3.490</td>
<td>2.460</td>
<td>0.408</td>
<td>0.318</td>
</tr>
<tr>
<td>06A BRE</td>
<td>0.142</td>
<td>2.860</td>
<td>1.760</td>
<td>0.300</td>
<td>0.331</td>
</tr>
<tr>
<td>06B BN</td>
<td>0.427</td>
<td>3.860</td>
<td>2.150</td>
<td>0.339</td>
<td>0.326</td>
</tr>
<tr>
<td>06D BN</td>
<td>0.573</td>
<td>3.540</td>
<td>2.160</td>
<td>0.338</td>
<td>0.339</td>
</tr>
<tr>
<td>071 BRE</td>
<td>0.417</td>
<td>3.210</td>
<td>1.870</td>
<td>0.306</td>
<td>0.348</td>
</tr>
<tr>
<td>077 LR</td>
<td>2.000</td>
<td>3.170</td>
<td>3.660</td>
<td>0.460</td>
<td>0.320</td>
</tr>
<tr>
<td>079 RE</td>
<td>0.658</td>
<td>2.860</td>
<td>2.170</td>
<td>0.396</td>
<td>0.333</td>
</tr>
<tr>
<td>07E BN</td>
<td>0.843</td>
<td>2.990</td>
<td>2.340</td>
<td>0.328</td>
<td>0.303</td>
</tr>
<tr>
<td>07F LR</td>
<td>0.420</td>
<td>2.850</td>
<td>1.950</td>
<td>0.332</td>
<td>0.309</td>
</tr>
<tr>
<td>083 LR</td>
<td>0.253</td>
<td>2.880</td>
<td>1.900</td>
<td>0.391</td>
<td>0.333</td>
</tr>
<tr>
<td>092 BRE</td>
<td>0.216</td>
<td>2.750</td>
<td>1.720</td>
<td>0.469</td>
<td>0.353</td>
</tr>
<tr>
<td>098 RE</td>
<td>0.187</td>
<td>2.830</td>
<td>1.720</td>
<td>0.360</td>
<td>0.339</td>
</tr>
<tr>
<td>09E BN</td>
<td>1.400</td>
<td>11.700</td>
<td>5.510</td>
<td>0.637</td>
<td>0.518</td>
</tr>
<tr>
<td>09F LR</td>
<td>0.062</td>
<td>2.920</td>
<td>1.660</td>
<td>0.309</td>
<td>0.334</td>
</tr>
<tr>
<td>0A0 RVCP</td>
<td>0.156</td>
<td>3.020</td>
<td>1.720</td>
<td>0.333</td>
<td>0.338</td>
</tr>
<tr>
<td>0A9 LVCP</td>
<td>0.012</td>
<td>3.600</td>
<td>1.660</td>
<td>0.293</td>
<td>0.311</td>
</tr>
<tr>
<td>0AA LR</td>
<td>−0.091</td>
<td>2.970</td>
<td>1.600</td>
<td>0.268</td>
<td>0.315</td>
</tr>
<tr>
<td>0B4 BN</td>
<td>0.154</td>
<td>4.280</td>
<td>1.870</td>
<td>0.305</td>
<td>0.338</td>
</tr>
<tr>
<td>0CA BN</td>
<td>−0.057</td>
<td>3.040</td>
<td>1.630</td>
<td>0.310</td>
<td>0.361</td>
</tr>
</tbody>
</table>

Notes: BMU and BLU are given as the averages over the phonation cycles per sample. The average values from the reference database of 52 non-dysphonic records to derive BMD, BLD, and BSD according to (13) are: $M_{b_0}=0.019$ gm, $R_{b_0}=12.36\text{ gm}.\text{sec}^{-1}$, and $K_{b_0}=2.24\times10^4\text{gm}.\text{sec}^{-2}$.

Abbreviations: BN, bilateral noduli; BP, bilateral polyp; BRE, bilateral Reinke’s edema; LR, larynx reflux; LVCP, left vocal cord polyp; RE, Reinke’s edema; RVCP, right vocal cord polyp.

aperture (AGA) and the cover dynamic component (CDC), related with the harmonics-to-noise ratio (HNR, Figure 16). Parameters \( p_{9-10} \) give the difference in the position and energy of the first spectral component in the cover-dynamics component relative to the average glottal aperture. The set of parameters \( p_{11-34} \) is defined to describe the spectral profile of the cover-dynamics component (Figure 17) as follows:

1. Parameters \( p_{11-14} \) estimate the spectral distribution of the cover-dynamics component divided into four energy bins.
2. Parameters \( p_{15-23} \) estimate the singular points on the profile of the spectral distribution of the cover-dynamics component such as the value of its origin, the amplitudes of the relative maxima and minima conforming to the first two v-grooved patterns detected, \(^{26}\) and the value at the highest frequency limit. (A v-grooved pattern is a minimum surrounded by two maxima in the shape of a “V.”)
3. Parameters \( p_{24-32} \) give the frequency positions of the singular values relative to the position of the first maximum.
4. Parameters \( p_{33-34} \) give a description of the sharpness of the first two v-grooved patterns found in the referred spectral distribution estimated as

\[
N_{nf} = \frac{T_m}{f_{21}} - 1; \quad T_m = T_n - T_{m1}; \quad f_{21} = \frac{f_{m2}}{f_{m1}} \tag{14}
\]

where \( T_m \) is the relative amplitude of the notch \( T_n \) with respect to the maximum at its left \( T_{m1} \) and \( f_{m1} \) and \( f_{m2} \) are the respective positions in frequency of the maxima to the left and right of the notch.

Figure 18 shows the values of the body mass, losses, and stiffness evaluated on a phonation cycle basis for the sample under study. Parameters \( p_{35-37} \) give the (interspeaker) deviations of the body mass, stiffness, and losses (BMD, BSD, and BLD) estimated from Equation 13 using average values of body mass, losses, and stiffness for each subject—over the complete phonation cycle set—against the average values \( M_{bs}, R_{bs}, \) and \( K_{bs} \) (given in the caption of Table 1) obtained from the cohort of 52 nondysphonic subjects used as reference. Parameters \( p_{38-40} \) give the BMU, BLU, and the BSU for the subject under study, as shown in Figure 19. Table 1 shows the pathologies detected in the set of the 20 dysphonic subjects (4 polyps, 6 vocal nodules, 5 Reinke’s edemae, and 5 pharyngo esophageal reflux-induced laryngitis and gives the respective values of the mass, losses, and stiffness deviations and the mass and losses unbalance for the cohort of 40 nondysphonic and dysphonic voices under study. It must be mentioned that samples “00B” and “024,” which were included \( a \ priori \) within the set of nondysphonic samples, when analyzed showed parameter deviations more in correspondence with mild dysphonic cases than with nondysphonic ones, as will be discussed in the sequel.

**RESULTS**

The parameter extraction process carried out on the cohort of 40 subjects under study (made of 20

---

**TABLE 2. List of 38 Parameters Used in Voice Analysis for the Case Under Study**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_2 )</td>
<td>Jitter</td>
</tr>
<tr>
<td>( p_{1.5} )</td>
<td>3 shimmer-related</td>
</tr>
<tr>
<td>( p_{6.7} )</td>
<td>2 glottal closure-related</td>
</tr>
<tr>
<td>( p_{8.10} )</td>
<td>3 HNR-related</td>
</tr>
<tr>
<td>( p_{11-14} )</td>
<td>4 mucosal wave psd in energy bins</td>
</tr>
<tr>
<td>( p_{15-23} )</td>
<td>9 mucosal wave psd singular point values</td>
</tr>
<tr>
<td>( p_{24-32} )</td>
<td>9 mucosal wave psd singular point positions</td>
</tr>
<tr>
<td>( p_{33-34} )</td>
<td>2 mucosal wave psd singularity profiles</td>
</tr>
<tr>
<td>( p_{35-37} )</td>
<td>3 biomechanical parameter deviations from Equation 13</td>
</tr>
<tr>
<td>( p_{38-39} )</td>
<td>2 biomechanical parameter unbalances from Equation 12</td>
</tr>
</tbody>
</table>

**Abbreviation:** psd, power spectral density.
dysphonic and 20 nondysphonic cases) produced a parameter matrix \( P = [p_{ij}] \) composed of \( m \) column vectors of \( r \) rows each. Row \( i \) contains the parameters corresponding to subject \( i \) (1 \( \leq i \leq r; r = 40 \)). Each 1 \( \leq j \leq m \) column corresponds to each specific parameter in Table 2, which is estimated for each subject in the cohort (in the case under study \( m = 38 \)). We will consider the parameter matrix \( P \) organized as a matrix of \( m \) column vectors of observations \( P = [p_1, \ldots, p_j, \ldots, p_m] \), each vector \( p_j \) collecting the observations corresponding to a different input parameter \( p_j \) over the cohort of subjects under analysis. Principal component analysis (PCA)\(^{35,36} \) is the specific methodology for the evaluation of the results. PCA allows compressing of the input space of observations to a new space of smaller dimensionality in terms of a new set of components, which collects most of the variance of the original set. In addition, the study of the relations between the original set of observation parameters and each principal component will allow us to infer which parameters of observations are more relevant for clustering. PCA consists of the linear transformation of the input matrix \( P = [p_1, \ldots, p_j, \ldots, p_m] \) of \( m \) parameters to a different matrix \( Y = [y_1, \ldots, y_j, \ldots, y_q] \) composed by \( q \) principal components. Both matrices \( P \) and \( Q \) describe \( r \) observations, each one corresponding to a voice sample in the cohort under study, but if the correlation among the parameters in the original set is large, \( Q \) will have a smaller dimensionality than \( P \); therefore, \( q << m \). In other words, PCA allows the compression of the redundant information present in the set of \( r = 40 \) voice traces with \( m = 38 \) parameters \( \{p_{2,39}\} \) accordingly with Table 2 to a set of \( r = 40 \) parameter sets but with a smaller number of principal parameters, which in our case can be reduced to \( q = 3 \), as it will be shown.

Two experiments were prepared for the analysis to establish the resolving capability of the set of parameters \( \{p_{2,39}\} \) in pathology detection tasks. One subset \( S_1 = \{p_{2,39}\} \) included the parameters

![FIGURE 13. Estimation of the pitch for the voice sample “00B” and its normalized value corresponding to parameter \( p_1 \).](image-url)
available except pitch ($p_1$) and body stiffness unbalance ($p_{40}$). (This parameter showed not to be relevant for the study, being highly correlated with BMU.) A second subset $S_2 = \{x_2, x_3, x_8, x_{35-39}\}$ included jitter, shimmer, HNR, deviations (BMD, BLD, and BSD), and unbalances (BMU and BLU). The selection of the parameters for this second set was derived from previous studies as these parameters were revealed to be among the most efficient ones in pathology detection. The raw matrices of data $P_{S1}$ (composed by the $m_1 = 38$ parameters defined in $S_1$ extracted from the $r = 40$ voice samples) and $P_{S2}$ (composed by $m_2 = 8$ parameters defined in $S_2$ extracted from the $r = 40$ voice samples) were divided into two clusters using hierarchical clustering. The results are given in Figure 20 as two bi-plots that give the relative position of each subject sample in terms of the two first principal components detected by PCA. The composition of the resulting clusters is given in Table 3. It may be observed that the clustering process using set $S_1$ assigned most nondysphonic subjects to cluster $c_{11}$ (with the exception of subjects 00B and 024), all dysphonic subjects being assigned to cluster $c_{12}$. In the case of using $S_2$, all nondysphonic subjects were clustered together in $c_{21}$, whereas all pathologic subjects were assigned to cluster $c_{22}$. Bi-plots in Figure 20 (right) show that cluster separation in this last case could be carried out using only the first two principal components, as observed from visual inspection—for instance, drawing a straight line through points $(-2,-2)$ and $(2,4)$.

Going one step further, PCA was used to establish the relative relevance of the original parameters in separating between nondysphonic and dysphonic cases. This study is based on a property of the covariance matrix of the input parameters $C = P'P$, providing information on the proportion in which each parameter in the input space $\{p_j\}$ is present in the components of the first principal component.

---

FIGURE 14. Estimations of parameter $p_2$ (normalized jitter), $p_6$ (normalized glottal source drift), and $p_7$ (normalized closure interval).
vector. It may be shown that the input parameters most contributing to the principal components would be those associated with the eigenvector components of C with largest absolute value. In this way it was determined that the three most relevant input parameters when using parameter set S₁ were, respectively, \( p_{37} \) (BSD), \( p_2 \) (jitter), and \( p_{36} \) (BLD). The 3-D plot showing the relative positions of each subject’s sample according to these three parameters is given in Figure 21. The study of this plot reveals that most samples clustered in \( c_{21} \) (encircled in full line) are associated with low values of BSD and BLD and with a small jitter, as opposed to samples clustered under \( c_{22} \), showing much larger values for the same parameters. Most samples in \( c_{22} \) are clustered in a central cloud (inside the dotted circle) with the exception of the especially anomalous cases “065,” “077,” and “09E,” which show the largest deviations. Cases of “00B” and “024” (inside the dash-dot circle) are not far away from the main grouping in \( c_{21} \), but according to \( S_1 \), they would have been classified as dysphonic because they show values for the BSD that double those of nondysphonic samples, although the evaluation protocol classified them \textit{a priori} as nondysphonic. This conclusion was confirmed by inspecting their respective values for the BSD given in Table 1, which were 1.25 and 1.2, respectively, or 225% and 220%, well above the deviations shown by nondysphonic samples. In general, it may be observed that large stiffness deviations are present in all pathological cases studied, which shows larger positive values than those in the nondysphonic cohort used as a reference. This result may indicate that looking for methods to efficiently and reliably estimate cord stiffness
could help in the automatic detection of pathology in voice.

**DISCUSSION**

The results produced by the current study suggest that a combination of classic acoustic parameters and biomechanical parameters extracted from the body-cover dynamics could improve voice evaluation methods with clinical purposes. The novelty of this approach relies on its capability for complementing classic distortion parameters derived from the raw signal of voice, with biomechanical parameters derived from glottal dynamics correlates (obliterating the interference of the vocal tract). This fact is especially relevant because biomechanical parameters may convey an important semantic load about the physiological state of the vocal cords. Using this approach, it has been possible to discriminate between nondysphonic and pathologic voices in this limited study. This is shown in Figure 21 and Table 3, where the sample cases were grouped into two separate clusters, clearly differentiating cases showing dysphonic behavior from those not showing it. The sensitivity of the differentiation allowed us to distinguish also those cases not showing a clear pathological condition (“00B” and “024”). It may not be surprising or coincidental that PCA highlighted BSD as the most relevant parameter for pathology determination, as it is well accepted that stiffness is associated with many voice pathologies. The fact that the second more relevant parameter seems to be jitter is also in agreement with classic studies in voice pathology research, as it is classically admitted. The third parameter in order of relevance as pointed out by PCA analysis is the BLD associated with the main peak profile (Q factor) in the glottal source power spectral density. This fact may not be causal, pointing to losses of energy due to nonelastic cord collision by forced (unnatural) phonation, with the unwanted consequences of cord heating, erosion of the cover, and damages to the associated tissues.

The case of the borderline sub-cluster formed by samples 00B and 024 was especially intriguing from the beginning of the study. In both cases, there was no evidence of vocal fold pathology as inferred from stroboscopic inspection even though auditory perceptual evaluation revealed moderate abnormal scorings. One possible explanation would be the presence of a subtle pathology undetectable by
stroboscopic means as these methods yield a zenithal (upright) perspective of the vocal fold surfaces, not revealing the inner deep subglottal structures of the vocal folds. This result means that the only pathologies appreciable using stroboscopic zenithal procedures will be those producing alterations of the viscoelastic properties of the cover (epithelium and superficial lamina propria) resulting in an abnormal pattern of vibration and mucosal wave propagation. The existence of other pathologies as, for example, those produced from an abnormal phonatory gesture would be hardly detectable by zenithal stroboscopy.

CONCLUSIONS

The methodology being exposed in this work is still under development, many of its fundamentals and implementation details being subject to revision and refinement. In this sense, there is a need for a detailed study of the inverse problems found in vocal cord biomechanical parameter estimation under a formal point of view, addressing concepts such as the uniqueness of the solutions and the calibration of the results. It requires collaborative research among the fields of larynx modeling and direct vocal fold dynamical imagery. The objective will be to obtain contrastive estimations of the vocal fold biomechanics from video-kimography (or a similar methodology) and the inversion of the glottal source from the voice trace, to determine the validity of studies like this presented. Another important objective to be covered is the statistical analysis of the biomechanical parameters extracted from a large population database of voice samples, including different pathological cases as well as normal ones, taking into account side conditions such as sex, age, and habits.

Leaving these objectives as open lines for future research, what can be concluded from this study on
the determination and parameterization of glottal dynamics correlates from the voice trace is that:

1. It seems feasible to establish a definition for the mucosal wave correlate based on considerations from vocal cord dynamics supported by three-mass models, decomposed as one-mass models for the body dynamics and two-mass models for the cover dynamics.
2. The extraction of the mucosal wave correlate or cover-dynamics component may be carried out using inverse filtering (Alku’s method) to obtain the glottal source correlate, after removing the body-dynamics component.
3. The power spectral distribution of the body- and cover-dynamics components seems to be directly related with specific system functions of the vocal cord dynamical structure.
4. The inversion of these dynamical system functions may help in obtaining correlates of the biomechanical parameters involved in the structural definition of the system, such as distributions of mass, losses, and stiffness.
5. The use of biomechanical parameter deviations from the averages defined over a nondysphonic set of voice records may be combined with classic distortion measurements in separating pathologic from dysphonic subjects.
6. Specific parameters related with the statistical deviation of stiffness and losses seem to play an important role in the separation of nondysphonic cases from dysphonic ones, showing high sensitivity in identifying boundary cases as well.
7. The selection of the parameters to optimally describe the presence of pathology remains an open issue. The combination of classic distortion measurements with biomechanical parameters and their extraction from the cover-dynamics component seem to be the key to better results.

FIGURE 18. Estimations of the body mass, stiffness, and losses as derived from the spectral profile of the body-dynamics component following Equations 9–11 for each phonation cycle of the voice sample “00B.”
Hierarchical k-means clustering and PCA have been revealed as powerful tools for pathology detection and classification, and in selecting the most suitable distortion parameters.

The methodology proposed is intended to fulfill desirable requirements of voice evaluation methods, such as objectivity, noninvasiveness, reliability, low cost, and low time consuming, favoring its possible application in screening programs. It is

### TABLE 3. Classification Results for Parameter Sets $S_1$ and $S_2$

<table>
<thead>
<tr>
<th>Set</th>
<th>Cluster</th>
<th>True Negatives</th>
<th>False Negatives</th>
<th>True Positives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>$c_{11}$ (o)</td>
<td>001, 003, 005, 007, 00A, 00E, 010, 018, 01C, 029, 02C, 02D, 032, 035 043, 047, 049, 04A</td>
<td></td>
<td></td>
<td>065, 069, 06A, 06B, 06D, 071, 077, 00B, 024</td>
</tr>
<tr>
<td></td>
<td>$c_{12}$ (〇)</td>
<td></td>
<td></td>
<td></td>
<td>079, 07E, 07F, 083, 092, 098, 09E, 09F, 0A0, 0A9, 0AA, 0B4, 0CA</td>
</tr>
<tr>
<td>$S_2$</td>
<td>$c_{21}$ (o)</td>
<td>001, 003, 005, 007, 00A, 00E, 010, 018, 01C, 029, 02C, 02D, 032, 035 043, 047, 049, 04A</td>
<td></td>
<td></td>
<td>00B, 024, 065, 069, 06A, 06B, 06D, 071, 077, 079, 07E, 07F, 083, 092, 098, 09E, 09F, 0A0, 0A9, 0AA, 0B4, 0CA</td>
</tr>
<tr>
<td></td>
<td>$c_{22}$ (〇)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 20. Biplots after dividing the input set of voice traces into two clusters with the whole set of parameters $S_1 = \{p_2-39\}$ (right) or the reduced one $S_2 = \{p_{23}, p_8, p_{35-39}\}$ (right). It may be observed that in this last case full separation of pathologic from non-pathologic cases can be easily accomplished by a straight line, such as the one drawn on the plot.

FIGURE 21. 3-D plot of clustering results in terms of the three most relevant observation parameters with respect to principal component analysis using subset $S_2$. Observe that samples “00B” and “024” (dash-dot line) have been located between the left-most subcluster (full line) and the main grouping of pathological cases (dotted line); the same is true with the estimations for the body stiffness deviations (absolute) with respect to the set of normal subjects used as a reference.
expected that this line of research could cover specific demands of clinical practice and provide future hints of applied research on voice dynamics.

REFERENCES
10. The Voice Center of Eastern Virginia Medical School. Available at: http://www.voice-center.com/larynx_ca.html

