

ISSUES IN SPEECH PRODUCTION BIOMECHANICS

*Comunicación efectuada por el Dr. Jorge Carlos Lucero
en la Academia Nacional de Ciencias de Buenos Aires
en la sesión privada extraordinaria
del 3 de marzo de 2008*

Abstract

This article reports our progress in the study of the biomechanics of speech production. Results obtained from the mathematical modeling of the vocal fold oscillation are shown, using various low dimensional models under several configurations, and their implications for speech production by children, men and women are discussed. Finally, our research on the analysis of articulatory kinematics and on the computer generation of facial animation of speech is briefly presented.

Resumen

Este artículo relata nuestros progresos en el estudio de la biomecánica de la producción del habla. Resultados obtenidos a partir del modelado matemático de la oscilación de las cuerdas vocales son mostrados, usando varios modelos de bajas dimensiones sobre diversas configuraciones, y sus implicaciones para la producción del habla por niños, hombres y mujeres son discutidas. Finalmente, nuestras investigaciones sobre el análisis de la cinemática de la articulación y sobre la generación de animaciones faciales del habla por computador son brevemente presentadas.

Introduction

The general purpose of our research is the development of mathematical and computational models of the organs involved in the production of speech, in order to simulate their dynamical behavior and understand their underlying physiology and control. At the same time, we work on the development of numerical algorithms to extract information from speech biomechanical signals, assess voice quality, and infer aspects of speech motor control. A number of potential applications may be considered, ranging from the audiovisual synthesis of speech by computer, to the development of computational tools for psychology and clinical studies.

In the next sections, we report and discuss our progress on the mathematical modeling of the vocal fold oscillation, and briefly present our research on related subjects of speech production.

The vocal fold oscillation

The mucosal wave model

The vocal folds at the larynx constitute an aeroelastic oscillator that acts as a sound source in voice production. Under proper conditions, the airflow blowing through the glottis induces their oscillation. The oscillation, in turn, modulates the airflow which, after going through and being modified by the oral and nasal cavities, produces the pressure wave that we perceive as voice.

Almost two decades ago, Titze (1988) set forth the dynamical principles of oscillation. He proposed a mucosal wave model in which motion of the vocal fold tissues is represented as a surface wave propagating in the direction of the airflow (Fig. 1).

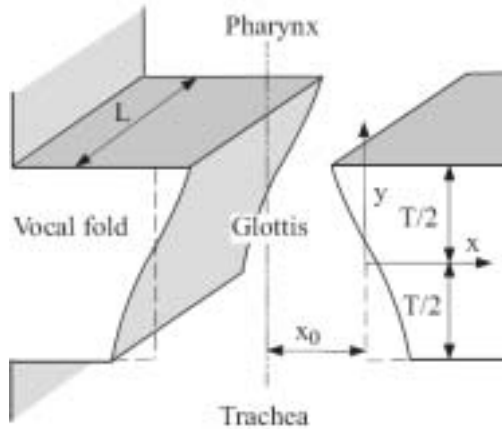


Figura 1. *Mucosal wave model of the vocal folds* (Lucero, 2005).

Let x be the displacement of the tissues from their rest position, at the midpoint of the glottis. Assuming a small time delay t for the mucosal wave to travel along the glottis, so that the approximation $x(t - \tau) \approx x(t) - \tau \dot{x}(t)$ holds, and lumping the mechanical properties of the tissues at the midpoint of the glottis, the equation of motion is

$$M\ddot{x} + B\dot{x} + Kx = \frac{2P_L \tau \dot{x}}{k_t(x_0 + x + \tau \dot{x})} \quad (1)$$

where M , B , and K are the mass, damping and stiffness coefficients per unit area of the medial surface, P_L is the lung pressure, x_0 is the glottal half-width at the rest position, and k_t is a transglottal pressure coefficient. When the lung pressure is small, the rest position of the vocal folds at $x = 0$ is stable, and hence no oscillation is possible. When it reaches the threshold value $P_L = k_t x_0 B / (2\tau)$, a Hopf bifurcation turns the rest position unstable and produces an oscillation. At the oscillation threshold, the energy transferred from the airflow to the vocal folds is large enough to overcome energy losses by dissipation at the tissues and may fuel an oscillatory motion of growing amplitude (Lucero, 2005).

Several extensions of the above model have been proposed in recent years. Laje et al. (2001) considered a nonlinear damping term of the form $B(1 + \eta x^2)\dot{x}$, where η is a phenomenological coefficient. We have shown that, according to the value of η , this model admits Hopf bifurcations of both the supercritical and subcritical type (Lucero, 2005). Further, inclusion of higher powers of x in the nonlinear damping term causes an oscillation hysteresis phenomenon for the onset-offset of the oscillation. This phenomenon has been observed in several experimental studies of phonation, and is critical to speech production because it determines patterns of voicing onset-offset during running speech (Lucero and Koenig, 2005a).

We have also lifted the restriction of a small value for the delay parameter τ , and consider the more general functional differential equation (Lucero and Koenig, 2007)

$$M\ddot{x} + B\dot{x} + Kx = \frac{P_L}{k_t} \frac{x(t + \tau) - x(t - \tau)}{x_0 + x(t + \tau)} \quad (2)$$

In this case, the oscillation threshold value of the lung pressure becomes $P_L = k_t x_0 B \omega \tau / (2\tau \sin \omega \tau)$. This extended equation contains the oscillation frequency ω explicitly, which was missing in the previous model by Titze. The threshold pressure increases with the oscillation frequency, in agreement with the experimental evidence.

The two-mass model

Another useful model of the vocal folds is the two-mass model (Ishizaka and Flanagan, 1972), in which each vocal fold is represented by two coupled mass-damper-spring systems (Fig. 2). Its equations of motion take the form

$$\begin{cases} m_1 \ddot{x}_1 + r_1 \dot{x}_1 + k_1 x_1 + k_c (x_1 - x_2) = f_1(x_1, x_2) \\ m_2 \ddot{x}_2 + r_2 \dot{x}_2 + k_2 x_2 + k_c (x_2 - x_1) = f_2(x_1, x_2) \end{cases} \quad (3)$$

where m_i , r_i , k_i ($i = 1,2$) are the mass, damping and stiffness coefficients of the masses, x_i is their displacement, and f_i are the forces exerted by the airflow. This model may be easily coupled to some representation of the upper and lower vocal tract systems, and is therefore more suited for detailed simulations of voice production.

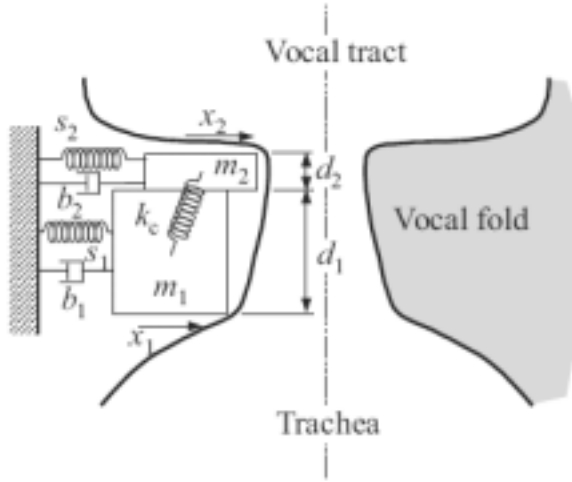


Figure 2: *The two-mass model of the vocal folds* (Lucero and Koenig, 2005a).

We have fitted the model to airflow data of male and female subjects producing the utterance /aha/ (Fig. 3), to investigate laryngeal control of voicing offset-onset (Lucero and Koenig, 2005a). Our results show that in both male and female speakers, devoicing is achieved by the combined action of the three parameters: vocal fold abduction, decrease of subglottal pressure, and an increase of vocal fold tension. Each of these actions has the effect of inhibiting the vocal fold oscillation. The results also show a large variability between speakers, which indicates that individual speakers develop unique laryngeal control strategies in running speech.

To explore further the relation of vocal fold oscillation conditions with laryngeal size, and investigate the development of laryngeal control with age, we have extended the two-mass model to a scalable version (Lucero and Koenig, 2005b). A single scaling factor is adopted

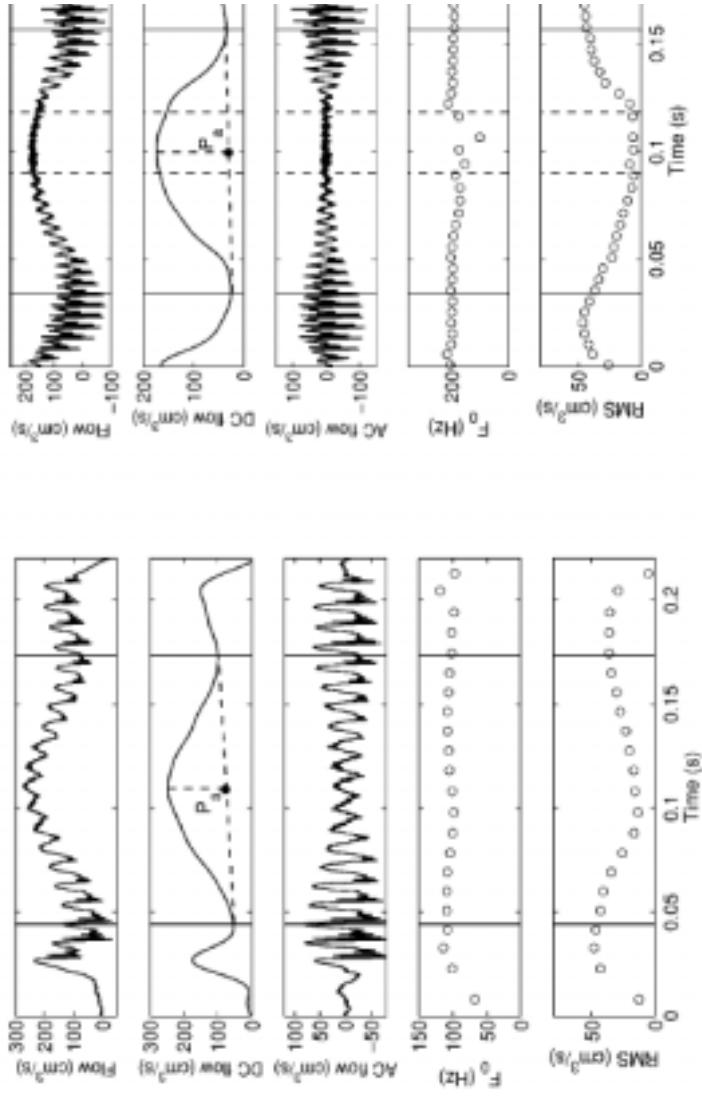


Figure 3. Airflow data from a male (left) and a female (right) subject (Lucero and Koenig, 2005a). From top to bottom: recorded airflow, dc component, ac component, fundamental frequency F_0 , and rms value of ac amplitude.

for all dimensions, and the various biomechanical parameters are scaled accordingly by assuming that the tissue elasticity modulus and the damping ratio are independent of size. Those are convenient simplifications of actual variations of size and tissue composition among men, women, and children, adopted for simplicity of the analysis.

The results show that the oscillation conditions of the vocal folds become more restricted as laryngeal size is reduced. This restriction results from a reduction of the glottal area in contact with the airflow, where the energy is transferred from the flow to the tissues. Thus, women will have more restricted conditions for vocal fold oscillation, which would explain their larger occurrence of devoicing in glottal abduction-adduction gestures (Koenig, 2000). Similarly, child vocal folds will have more restricted oscillation conditions than adults, which could result in their higher values of subglottal pressures observed during speech (e.g., Stathopoulos and Sapienza, 1993).

The values that main control parameters of the model must meet in order to produce the oscillation are related by the equation $\beta P_L = k a_0 Q$, where β is the scaling factor, k is constant, a_0 is the prephonatory glottal area, and Q is a vocal fold tension parameter. The equation suggests again the possibility of different strategies for speakers to achieve similar patterns of voicing and devoicing during speech: speakers within a size group could adopt different vocal tract postures and adjust accordingly the vocal fold tension to arrive at the same phonatory result.

Conclusions

Our research shows, in general, that low-dimensional mathematical models of the vocal folds, in spite of their simplicity, are capable of reproducing their dynamical behavior at phonation with good detail, and with control parameters of physiologically realistic values. Fitting models to speech data is an important technique for making quantitative predictions of speech biomechanics. Most of our work is based on oral airflow signals, which are relatively easy to obtain, and provide a convenient noninvasive means of studying laryngeal behavior. This is particularly important when studying the development of speech motor control in children.

Besides the vocal folds, our group also carries out research on the analysis of articulatory kinematics during speech. Specifically, we

develop computational algorithms based on Functional Data Analysis (Ramsay and Silverman, 1997) to extract patterns and assess variability of articulator trajectories (tongue, lips, jaw) across repetitions. It is known that variability is larger in children than in adults, and also in speakers with neurological disorders compared to normal subjects. Our results show that variability varies as a function of the phonetic requirements of the speech task and the biomechanical constraints imposed by the articulatory structure involved (Lucero and Löfqvist, 2005). A potential area of clinical application is in the assessment of the efficacy of treatment, through the analysis of profiles of an individual speaker's production variability.

A third area of interest is the development of 3D models of human face biomechanics for the production of computer-generated animations of speech. We are following an empirical modeling strategy, in which mathematical relations between facial kinematic parameters measured during speech are constructed. We have proposed an algorithm based on the QR decomposition of linear algebra, which identifies facial regions of independent motion (Lucero et al, 2005). The regions represent the degrees of freedom of the system, and are determined by the patterns of muscle contractions and the biophysical characteristics of skin tissue. They vary for different subjects, and to some degree, for different data sets. In all studied cases, however, the lower lip and both lip corners tend to be among the most independent kinematic regions.

For more information on the subjects reported in this article, we refer the reader to the cited bibliography, and to our web page in <http://www.mat.unb.br/lucero/>.

References

- Ishizaka, K., and Flanagan, J. L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal folds", *Bell Syst. Tech. J.* 51, 1233 – 1268.
- Koenig, L. L. (2000). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds", *J. Speech Lang. Hear. Res.* 43, 121 – 1228.
- Laje, R., Gardner, T. J. and Mindlin, G. B. (2001). "Continuous model for vocal fold oscillations to study the effect of feedback", *Phys. Rev. E* 64, 056201.
- Lucero, J. C. (2005). "Bifurcations and limit cycles in a model for a vocal fold oscillator", *Comm. Math. Sci.* 3, 517 – 529.

- Lucero, J. C. and Koenig, L. L. (2005a). "Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control", *J. Acoust. Soc. Am.* 117, 1362 – 1372.
- Lucero, J. C. and Koenig, L. L. (2005b). "Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds", *J. Acoust. Soc. Am.* 118, 2798 – 2801.
- Lucero, J. C. and Koenig, L. L. (2007). "On the relation between the phonation threshold lung pressure and the oscillation frequency of the vocal folds", *J. Acoust. Soc. Am.* 121, 3280 – 3283.
- Lucero, J. C. and Löfqvist, A. (2005). "Measures of articulatory variability in VCV sequences", *Acoustics Research Letters Online* 6, 80 – 84.
- Lucero, J. C., Maciel, S. T. R., Johns, D. A. and Munhall, K. G. (2005). "Empirical modeling of human face kinematics during speech using motion clustering", *J. Acoust. Soc. Am.* 118, 405 – 409.
- Ramsay, J. O. and Silverman, B.W. (1997). *Functional Data Analysis* (Springer-Verlag, NewYork)
- Stathopoulos, E. T. and Sapienza, C. (1993). "Respiratory and laryngeal measures of children during vocal intensity variation", *J. Acoust. Soc. Am.* 94, 2531 – 2543.
- Titze, I. R. (1988). "The physics of small-amplitude oscillation of the vocal folds", *J. Acoust. Soc. Am.* 83, 1536 – 1552.

Department of Mathematics, University of Brasilia
 Brasilia DF 70910-900, Brazil