Simulations of VhV sequences in children

Jorge C. Lucero* and **Laura L. Koenig**†

* Department of Mathematics, University of Brasilia, Brazil
lucero@mat.unb.br

† Haskins Laboratories & Long Island University, Brooklyn Campus, USA
koenig@haskins.yale.edu

ABSTRACT

We explore the adequacy of low-dimensional models of the larynx to reproduce aerodynamic patterns of child speakers in the vicinity of an abduction gesture. A model of the child larynx is created by scaling an adult two-mass model coupled to a two-tube approximation of the vocal tract. This model next is fitted to oral airflow data of VhV sequences recorded from 5-year-old speakers, and using the subglottal pressure, Q-factor, and glottal width as control parameters. Optimal values of the control parameters are computed by using the values of DC flow, rms AC flow, and f_0 measured at the /h/ abduction and flanking vowels as targets. The results show that the model is capable of reproducing the aerodynamic patterns with good approximation, allowing us to infer some aspects of voice control in individual speakers.

1 INTRODUCTION

Our goal in this work is to use laryngeal modeling to learn more about how the vibratory behavior of a child's larynx may differ from that of an adult. Specifically, we extend previous work, which used low-dimensional laryngeal models to reproduce voicing behavior in men and women, to consider analogous signals collected from 5-year-old children. Past research has indicated that low-dimensional models, though simpler than the human larynx, still capture many significant aspects of vocal fold vibration for speech. In our own work on adults, for example, a modified two-mass model coupled to an simple (two-tube) upper vocal tract reproduces hysteresis effects in voicing offset and onset around an abduction gesture, as well as f_0 and pulse amplitude variation near voicing thresholds [1].

Modeling is likely to be particularly useful for understanding the characteristics of the developing larynx. Anatomical considerations indicate that tissue characteristics may differ considerably between children in adults; in particular, relatively shorter membranous vocal folds in children should yield higher overall vocal fold stiffness than in adults [2]. Quantifying such effects and even making precise measurements of dimensions in the real child larynx is difficult for many reasons, however: Dissection studies suffer from low sample sizes, and the changes in mechanical properties that occur in excised tissues make estimates of physical characteristics such as stiffness questionable. Determining laryngeal dimensions from anatomical imaging is complicated by problems of obtaining clear boundaries of cartilage and soft tissues, especially when the structures in question undergo movement during normal functioning. Finally, direct measurement of the larynx is invasive and usually not an option for studying healthy children.

In addition to differences in overall dimensions and tissue characteristics, physiological differences between children and adults may affect children's phonatory behavior. In particular, children generally have higher subglottal driving pressures than adults for comparable speaking tasks, a finding usually attributed to higher resistances in the lower airways [3]. Modeling permits us to vary anatomical and physiological parameters independently, to determine how each affects vibratory behavior.

2 DATA

We collected data from 2 normally-developing English speaking 5-year-old children. The data consisted of oral airflow signals, recorded using a Rothenberg mask while the subjects were repeatedly producing the utterance "Papa Hopper", with stress on the third syllable. The records were filtered at 5 kHz, and sampled at 10 kHz. For each subject, 6 typical records were selected through visual inspection, in 3 of which the /h/ was devoiced, and in the remaining 3 the /h/ was fully voiced. In each of these records, the portion corresponding to the production of utterance /aha/ was extracted for further processing with Matlab software.

The DC component of the records was obtained by low-pass filtering them at a 50 Hz cut-off, to eliminate glottal pulses. Next, the AC component was computed

ISBN 1-876346-48-5 © 2003 UAB

as the difference between the unfiltered original record and its DC component. From the AC component, we next computed its fundamental frequency f_0 and rms amplitude, using a zero crossing algorithm with low pass filtering [4].

Figure 1 shows a data example for subject 5F1, in a case of unvoiced /h/. At the top, we have the recorded flow signal, for the period corresponding to the utterance /aha/. Below, we have the computed DC and AC components. Let us recall that airflow is roughly proportional to glottal area, when other parameters are constant. Hence, the curve of the DC flow may be considered as an approximate measure of the glottal area, clearly showing the abduction-adduction gesture for the production of /h/. The plots also show that voicing stops almost at the peak of the glottal abduction, and restarts near the end of the subsequent adduction, in a clear hysteresis effect: to restart voice, the vocal folds must be driven closer together than the position at which voice stopped. At the bottom, we have plots for f_0, and the rms AC amplitude, computed for each cycle of the AC flow. The region with no marks corresponds to the unvoiced period.

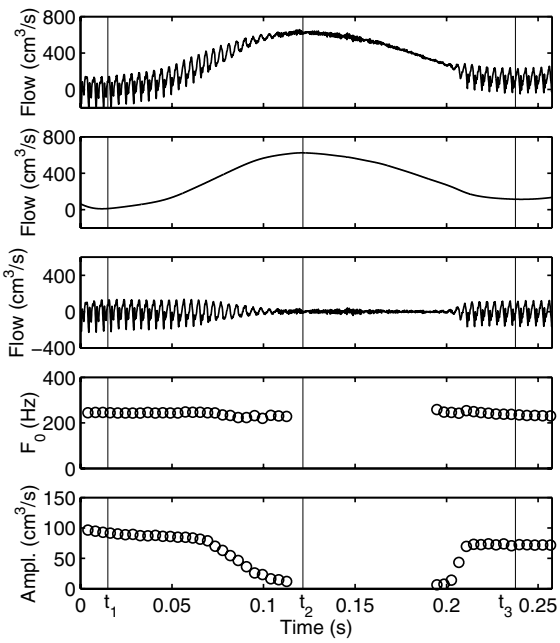

Figure 1: Data example for subject 5F1, in a case of unvoiced /h/. From top to bottom: recorded oral airflow, DC flow, AC flow, fundamental frequency, and rms amplitude of AC flow. Marks t_1, t_2, and t_3 are the times defined as targets for fitting the laryngeal models

Figure 2 shows a data example for the other subject, 5F3, in a case of voiced /h/. There is a glottal abduction and a decrease of the AC amplitude, but voice never stops.

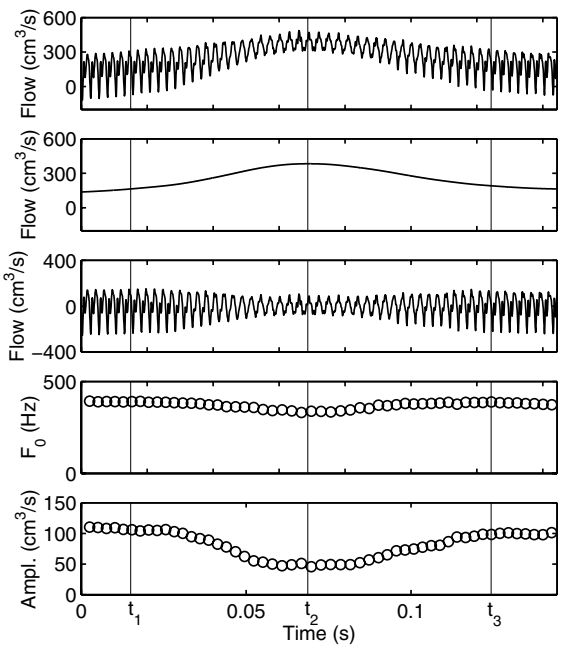

Figure 2: Data example for subject 5F3, in a case of voiced /h/. From top to bottom: recorded oral airflow, DC flow, AC flow, fundamental frequency, and rms amplitude of AC flow. Marks t_1, t_2, and t_3 are the times defined as targets for fitting the laryngeal models

3 MODEL

The complete model was implemented as a computer program using Matlab software.

For the larynx, we used a version of the popular two-mass model by Ishizaka and Flanagan [5], with a non-linear characteristic for the tissue viscosity [1]. The vocal tract was modeled with a simple two-tube approximation, in configuration for vowel /a/ [6].

The standard configuration of the two-mass plus two-tube vocal tract approximation, which correspond to an adult male, was transformed to a 5-year-old configuration using a model of the developing vocal tract [7]. We calculated a single scaling factor of $\beta = 0.64$ (approximate relation of a 5-year-old vocal tract length to an adult male one), which was applied to all parameters of the model, as follows: letting d, k, m, ξ denote the various dimensions, stiffness coefficients, masses, and damping ratios of the model, and subindices M and K denote male adult and children values, then

$$\begin{aligned} d_K &= \beta d_M \\ k_K &= (1/\beta) k_M \\ m_K &= \beta^3 m_M \\ \xi_K &= \xi_M \end{aligned} \tag{1}$$

The lengths of the tubes in the vocal tract model (oral cavity and pharynx) were also adjusted to match the length relation of a 5-year-old (64.3 mm and 42.7 mm,

respectively [7]). As control parameters of the model (to be determined optimally), we selected the neutral glottal width x_0, Q factor (scaling coefficient for the natural frequencies of the model [5]), and subglottal pressure P_s.

As targets for fitting the model, we used three parameters: fundamental frequency f_0, DC flow and rms AC flow, computed from the measured data at three instants of time : t_1, at the beginning of the glottal abduction, t_2, at the peak of the /h/ abduction, and t_3, at the end of the glottal adduction (see Figs. 1 and 2). These three instants were determined by visual inspection of the recorded signals. Therefore, we determined the values of the control parameters at those three instants of time that produced the closest values of f_0, DC flow, and rms AC flow, to the targets. In periods $t \leq t_1$ and $t \geq t_3$, the control parameters were fixed at constant values, and in period $t_1 < t < t_3$ (i.e., the /h/ abduction-adduction gesture), we adopted sinusoidal variations for them.

The fitting was done by minimizing the total square error between the targets and the simulation results, and using the standard optimization algorithm provided in Matlab.

4 RESULTS

4.1 SIMULATIONS

Figs. 3 and 4 show simulations results after fitting the model to the data in Figs. 1 and 2, respectively. Comparing those figures, some differences between simulated results and recorded data are seen, particularly during the abduction and adduction. These difference are probably caused by the assumed sinusoidal variations of the control parameters for this region, which might be a poor approximation. However, we may say that in general, the model fits the data with good approximation.

All other cases for both subjects, with and without devoicing, produced similar results to those shown here.

4.2 CONTROL PARAMETERS

Figs. 5 shows the fitted curves of the control parameters for subject 5F1. The glottal width plot clearly shows the glottal abduction-adduction gesture for /h/. Its amplitude reaches values up to 0.1 cm, which are in the same range as those obtained for adult subjects [1]. The subglottal pressure is in the range 9.6–18 cm H_2O, which is about twice the values for adults. The Q factor is in the range 0.56–0.82. Recalling that the vocal fold stiffness has been scaled with a factor 1/0.64 from its adult value, then the child stiffness results are similar to the adult. Comparing next the voiced /h/ cases with the devoiced ones, we note that devoicing is produced by combination of a large increase in glottal opening, a decrease of subglottal pressure, and a slight increase of Q factor. Each of these actions have the effect of inhibiting the vocal fold vibration.

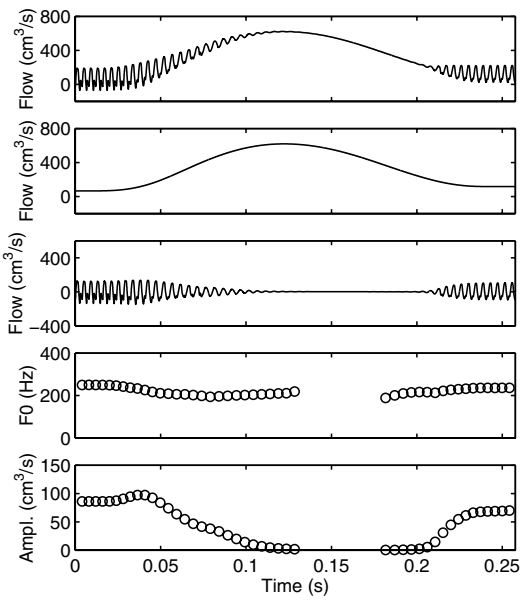

Figure 3: Simulations results using the data in Fig. 1. From top to bottom: oral airflow, DC flow, AC flow, fundamental frequency, and rms amplitude of AC flow.

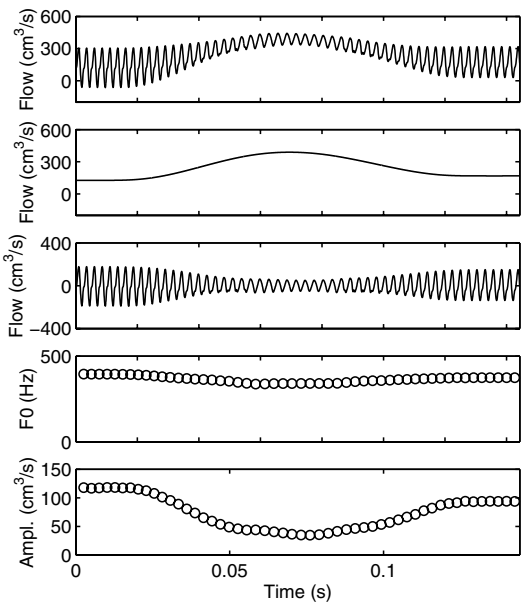

Figure 4: Simulations results using the data in Fig. 2. From top to bottom: oral airflow, DC flow, AC flow, fundamental frequency, and rms amplitude of AC flow.

Fig. 6 show the control parameters for subject 5F3. The glottal width and subglottal pressure values are in similar range as the previous subject. The Q factor is now in the range 0.70–1.24, which implies a vocal fold stiffness 1.09–1.93 higher than the standard adult

ISBN 1-876346-48-5 © 2003 UAB

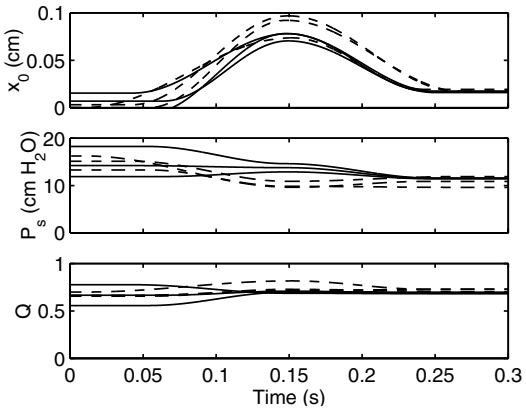

Figure 5: Control parameters for subject 5F1. Top: neutral glottal width. Middle: subglottal pressure. Bottom: Q factor. Full curves: voiced /h/. Broken curves: unvoiced /h/.

one. Comparing the voice vs. unvoiced cases, we note that the glottal abduction has similar amplitudes in both cases. However, in the devoiced cases, the subglottal pressure is much lower. This lower value tends to inhibit the vocal fold vibration, and hence the glottal width does not need to reach a large level to stop voicing. In the unvoiced cases, there is also a slight increase of Q factor during the glottal abduction, which also helps to stop voice.

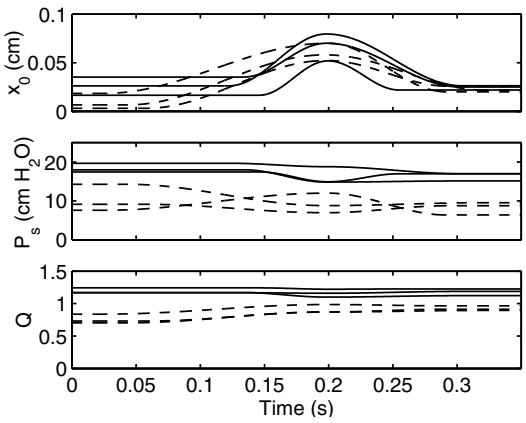

Figure 6: Control parameters for subject 5F3. Top: neutral glottal width. Middle: subglottal pressure. Bottom: Q factor. Full curves: voiced /h/. Broken curves: unvoiced /h/.

5 CONCLUSION

In general, the results show that the two-mass model plus two-tube vocal tract approximation seems capable of reproducing the aerodynamic patterns of child speakers. At the same time, some of the parameter values obtained via simulation apparently differ somewhat from those used by the child speakers recorded here. The glottal width values obtained, in the same range as those obtained for adults, may be too large for children, and the subglottal pressure values are higher than those recorded for the children here (8-10 cm H_20 for 5F1, and 10-14 cmH_20 for 5F3).

In future work, we plan to vary other characteristics of the model, including tissue characteristics, to determine how these affect the simulation results. We will also extend our work to a larger group of speakers, both child and adult, to identify patterns and possible differences in voicing control as a function of age.

ACKNOWLEDGMENTS

These work was supported by grants of the NIH to Haskins Laboratories, and CNPq (Brazil).

REFERENCES

[1] J. C. Lucero and L. L. Koenig, "Simulations of voice onset-offset hysteresis in /aha/ utterances," *Journal of the Acoustical Society of America*, vol. 108, pp. 2508, 2000.

[2] S. Kurita M. Hirano and T. Nakashima, "Growth, development, and aging of human vocal folds," in *Vocal fold physiology: Contemporary research and clinical issues*, D. M. Bless and J. H. Abbs, Eds., pp. 22–43. San Diego: College Hill Press, 1983.

[3] E. Stathopoulos and C. Sapienza, "Developmental changes in laryngeal and respiratory function with variations in sound pressure level," *J. Speech Lan. Hear. Res.*, vol. 40, pp. 595–614, 1997.

[4] I. R. Titze and H. Liang, "Comparison of f_0 extraction methods for high-precision voice perturbation measurements," *Journal of Speech and Hearing Research*, vol. 36, pp. 1120–1133, 1993.

[5] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Systems Technical Journal*, vol. 51, pp. 1233–1268, 1972.

[6] I. R. Titze, *Principles of Voice Production*, Englewood Cliffs NJ: Prentice-Hall, 1993.

[7] U. Goldstein, *An articulatory model for the vocal tracts of growing children*, doctoral dissertation, MIT, 1980.