A Discrete-Time Aerodynamic Model of the Vocal Tract

Ronald L. Sprouse

Department of Linguistics, University of California, Berkeley
E-mail: `ronald@berkeley.edu`

Abstract

Computer simulations of speech processes can lead to better understanding of real-world speech phenomena and can be used to predict and interpret the results of speech production experiments. An aerodynamic model that updates vocal tract parameters over small time steps is presented. Some initial results illustrating how this model (implemented in Matlab) works demonstrate how changes in articulation can lead to differences in voicing and frication.

1 Introduction

This paper reports on a discrete-time aerodynamic model of the vocal tract that updates, refines, and elaborates the model described by Ohala [1], in which the vocal tract is represented as two serially-connected cavities—the subglottal and supraglottal volumes—and speech processes are initiated and maintained by the elastic recoil force of the chest walls. Variations in pressure, flow, and related variables (e.g. particle velocity) are taken to be a function of the resistance presented by the glottal and the oral articulators. Other models with a similar general approach but different implementation include Rothenberg [2] and Westbury and Keating [3].

The present model improves on the Ohala model in that additional vocal tract parameters are introduced, allowing for the simulation of a greater variety of phonetic articulations. Specifically, the present model provides the following new parameters:

- **Glottal offset:** The glottal offset is used to simulate vertical movement of the glottis. Increasing the glottal offset during a stop closure is used for modeling ejective segments; a decrease in this parameter is used to model implosives and to model larynx-lowering as a strategy for maintaining voicing during stop closures [4].
- **Velar opening:** The velar opening is introduced to model the effects of nasal air flow;
- **Tissue compliance:** Tissue compliance is modeled as a dynamic response to the impinging pressures of the air in the vocal tract.

The model can be used to simulate experimental conditions in addition to natural speech. As an example, the velar opening parameter provides a means of simulating experiments that involve artificial venting of oral pressure, as in [5].

2 Implementation

The model is implemented in Matlab. Speech processes are simulated by calculating and successively updating airflow and volume changes of the vocal tract cavities over very small time steps, then low-pass filtered to remove unneeded high frequency components.

2.1 Input parameters

Model inputs are of two types—initial conditions that describe the overall conditions under which an utterance is made, and articulatory gestures that are under the control of the model speaker. The starting conditions include:

- the initial subglottal and supraglottal volumes and air masses;
- the elastic recoil force of the chest walls;
- physical constants, such as atmospheric pressure (which is assumed to be constant for the duration of the modeled speech event).

Speaker-initiated events are encoded as the time-varying inputs to the system. These include:

- changes of the model speaker's target size for the glottal, oral, and velar apertures over time;
- the front/back movement of the location of the greatest oral constriction in the oral cavity; and
- the vertical movement of the glottis up or down from a neutral position.

The recoil force of the chest walls initiates and maintains the speech process, which is influenced by events in the other vocal tract parameters. Any sequence of particular speech segments can be modeled by choosing appropriate articulatory events in the vocal tract. For example, the sequence [at^{h}i] can be modeled by specifying events that operate on the glottal aperture, changing it from a partially-constricted setting for voicing during [a] to an open setting for the voiceless stop closure and the aspirated sequence, and returning it to the voicing posture for [i]; these events overlap with oral aperture events that change the oral articulation from an open articulation during [a] to fully-closed during the stop closure, and back to an open articulation for [i] (though less open than the articulation of [a]). By comparison, the unaspirated sequence [ati] is configured by a minor change in the relative timing of the glottal and oral events such that the glottal aperture returns to the voicing setting concurrent with the release of the stop closure.

2.1 Outputs

Model outputs are in the form of time-series vectors containing the internal model states. At each time step, the following values are calculated:

- volume of the subglottal and supraglottal cavities;
- air pressure of the subglottal and supraglottal cavities;
- airflow through the glottis and oral place of articulation;
- particle velocity of the air flowing through the glottis and oral place of articulation;
- whether voicing is present.

While the model does predict whether a sound is voiced or not at a particular point in time, based on the amount of pressure drop across the glottis, it does not attempt to directly model vocal fold vibrations.

3 Results

The ultimate goal will be to calibrate the model with natural speech data in order to produce quantitative predictions about speech events. In its present state the model does not provide outputs in physically-interpretable units, but it does generate a variety of articulations that can be used to make qualitative pairwise comparisons. Some model outputs will be presented: the effects of nasal leakage and larynx lowering on voicing during stop closure; the effect of vowel height on oral pressure and particle velocity, hence on voicing and frication.

3.1 Nasal leakage

Figures 1 and 2 illustrate the effect of velar leakage on voicing. Figure 1 shows fully-oral [adi], and Figure 2 shows the same sequence of segments with a small amount of velar leakage, showing how velar leakage can extend the duration of voicing during a stop closure. The top portions of each figure display the oral aperture inputs for [adi], along with the predicted voicing. The low part of the oral aperture line indicates full closure during the stop, and the higher parts indicate the relatively open apertures characteristic of the vowels (and the transitions between the vowels and stop). The dashed line in Figure 2 indicates the presence of velar leakage during most of the stop closure.

In the bottom part of Figure 1 oral pressure approaches subglottal pressure during the stop closure until the two are almost equal, and voicing cease when the difference in pressure is too small to sustain voicing. In Figure 2 the velar leakage during the stop closure prevents oral pressure from rising to subglottal pressure as rapidly as it would in the absence of the leak, and voicing is maintained throughout the stop closure.

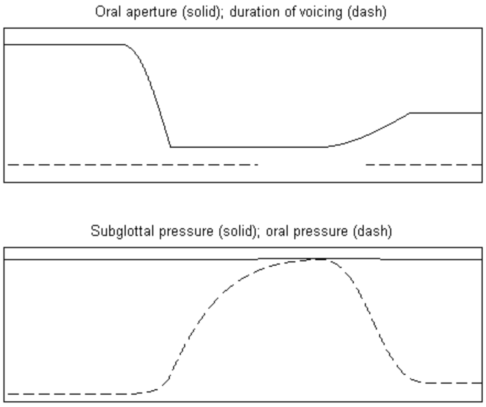

Figure 1: *[adi]*

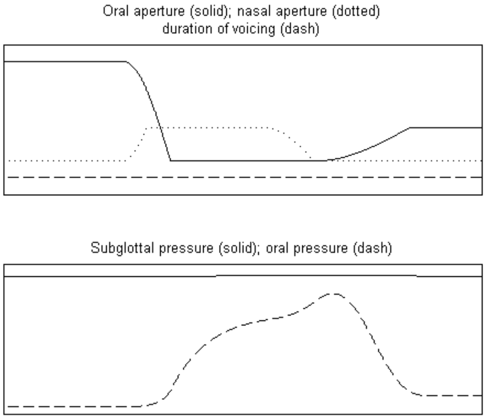

Figure 2: *[adi] with velar leakage*

3.2 Larynx lowering

Figures 3 and 4 show how larynx lowering has an effect on voicing during a stop closure similar to the effect of velar leakage. Figure 3 shows [aba], with voicing extinguished during the stop closure. Figure 4 shows the same sequence with the supraglottal cavity enlarged by using the glottal offset parameter. The supraglottal cavity enlargement that results is shown by the dotted line. This enlargement depresses oral pressure, and voicing is sustained throughout the stop closure.

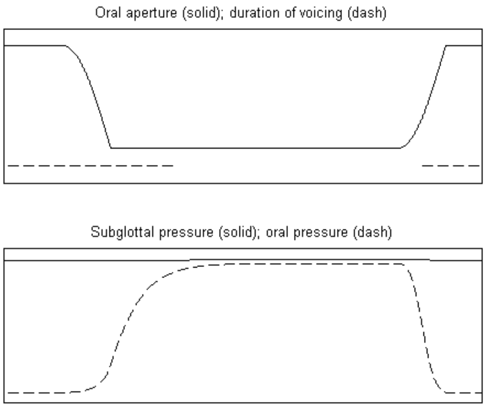

Figure 3: *[aba]*

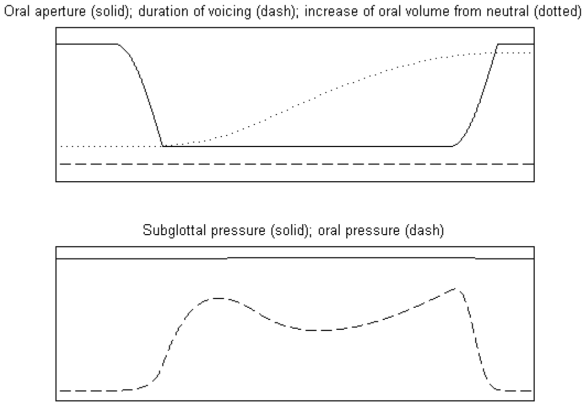

Figure 4: *[aba] with larynx lowering*

3.2 Vowel height

Figure 5 illustrates the effect of vowel height on oral pressure and oral velocity. High vowels tend to be subject to devoicing and/or frication more often than low vowels after a voiceless stop [6]. As the figure shows, the relatively smaller aperture of [i] compared to [a] (top panel) results in a slower decrease in oral pressure during a stop release. High oral pressure leads to a lower transglottal pressure difference, which tends to inhibit voicing, as already seen in Figures 1 and 3. Particle velocity is inversely proportion to the oral aperture (velocity = flow/area), and the smaller oral aperture of [i] therefore leads to a higher oral velocity, which is the source of noise that may be interpreted as frication.

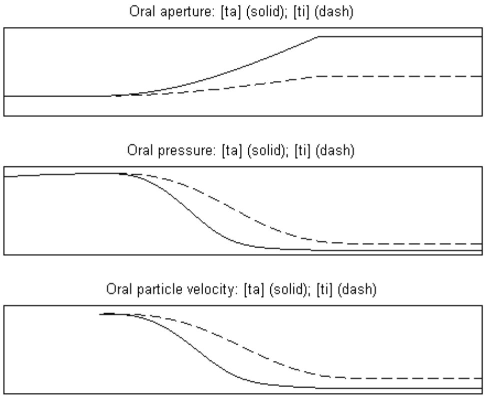

Figure 5: *release of [ta] and [ti]*

4 Conclusion

The presented results show that the discrete-time aerodynamic model may be used to make pairwise comparisons between similar but non-identical articulations, even in the absence of real-world output units. One advantage to using a computer simulation is that it provides a way of exploring physical parameters that are hard to measure, notably subglottal pressure. In addition, the presented model's straightforward implementation as a series of time steps does not require advanced mathematical skills and therefore provides a teaching tool that can be used to introduce concepts in speech aerodynamics in an easily understood way.

References

[1] J. J. Ohala. A mathematical model of speech aerodynamics. In: G. Fant (ed.), *Speech Communication*, [Proc., Speech Comm. Seminar. Stockholm, 1-3 Aug. 1974.] Vol. 2: Speech production and synthesis by rule. Almqvist & Wiksell, Stockholm., 1975.

[2] M. Rothenberg. The breath-stream dynamics of simple-released plosive production. *Bibliotheca Phonetica* 6. 1968.

[3] J. R. Westbury & P. A. Keating. On the naturalness of stop consonant voicing. *Journal of Linguistics* (22):145-166. 1986.

[4] G. N. Clements & S. Osu. Explosives, implosives and nonexplosives. In: C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology* 7. Mouton de Gruyter, Berlin, 2002.

[5] R. L. Sprouse, M. J. Solé & J. J. Ohala. Oral cavity enlargement in retroflex stops. This volume. 2008.

[6] J. Greenberg. Synchronic and diachronic universals in phonology. *Language* 42:508-17. 1966.