INTERACTION BETWEEN GLOTTAL AND VOCAL-TRACT AERODYNAMICS IN A COMPREHENSIVE MODEL OF THE SPEECH APPARATUS

Paul Boersma
Institute of Phonetic Sciences, Amsterdam, The Netherlands

ABSTRACT

Our computer model of the speech apparatus converts muscle activities into the resulting tissue movements, air pressures, air velocities, and sound.

THE MODEL

The entire speech apparatus (lungs, glottis, mouth, nose) is modelled with 80 tube sections that contain air (fig. 1).

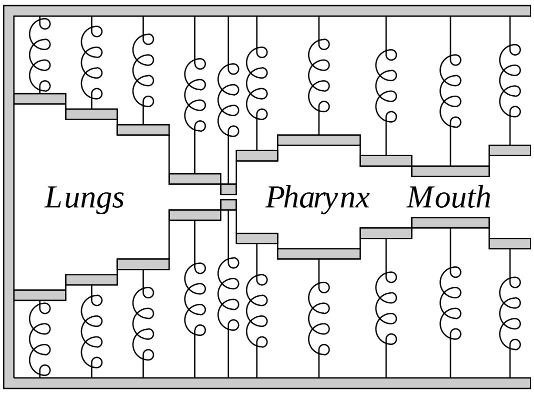


Figure 1. The whole speech apparatus, including the glottis, is modelled with the same kind of tubes. Only a few of these tubes are shown here.

Each tube section has (a) two parallel stiff walls, (b) two near-parallel walls that can move under aerodynamic and myoelastic forces and whose equilibrium positions and tensions are controlled by the muscles, and (c) two boundaries that connect the tube to the rest of the world.

Figure 2 shows the four different forms of these boundaries: they are either (1) closed, (2) connected to one other section, (3) connected to two other sections, or (4) open to the outer air.

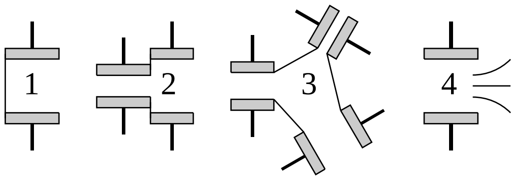


Figure 2. Types of tube boundaries.

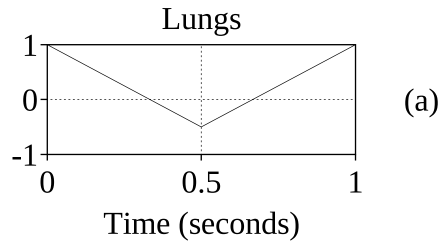
The sound-generating algorithm is not specific to speech: it works equally well for an arbitrary structure of ducts.

All tube sections can have different and time-varying widths and lengths. The finite-differencing integration scheme that solves the resulting aerodynamic problem without approximating away any pumping and sucking effects, was presented in [1]. In the present paper, I will show examples of some of the phenomena that the model can describe realistically.

Our model speaker is characterized as an average adult woman.

LUNG PRESSURE

Because the lungs are described in the same way as the vocal tract, subglottal pressure is a direct consequence of controlling the equilibrium (target) width of the lungs.



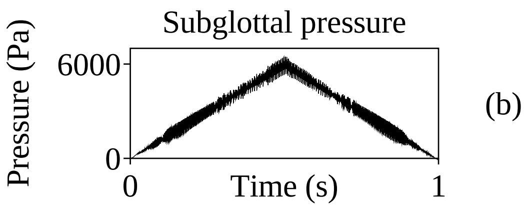


Figure 3. Lungs: input (target width) and output (subglottal pressure), if the vocal cords are adducted (phonation).

Figure 3 shows how this relation is realized during the production of the vowel [a]. The change in target width is much more reflected in the lung pressure than in the lung volume. The intensity of the uttered sound is an increasing function of lung pressure: the slope is 10 dB per kPa for pressures below 1.3 kPa, and 2 dB per kPa for pressures above. The fundamental frequency varies by 100 Hz and 20 Hz per kPa, respectively.

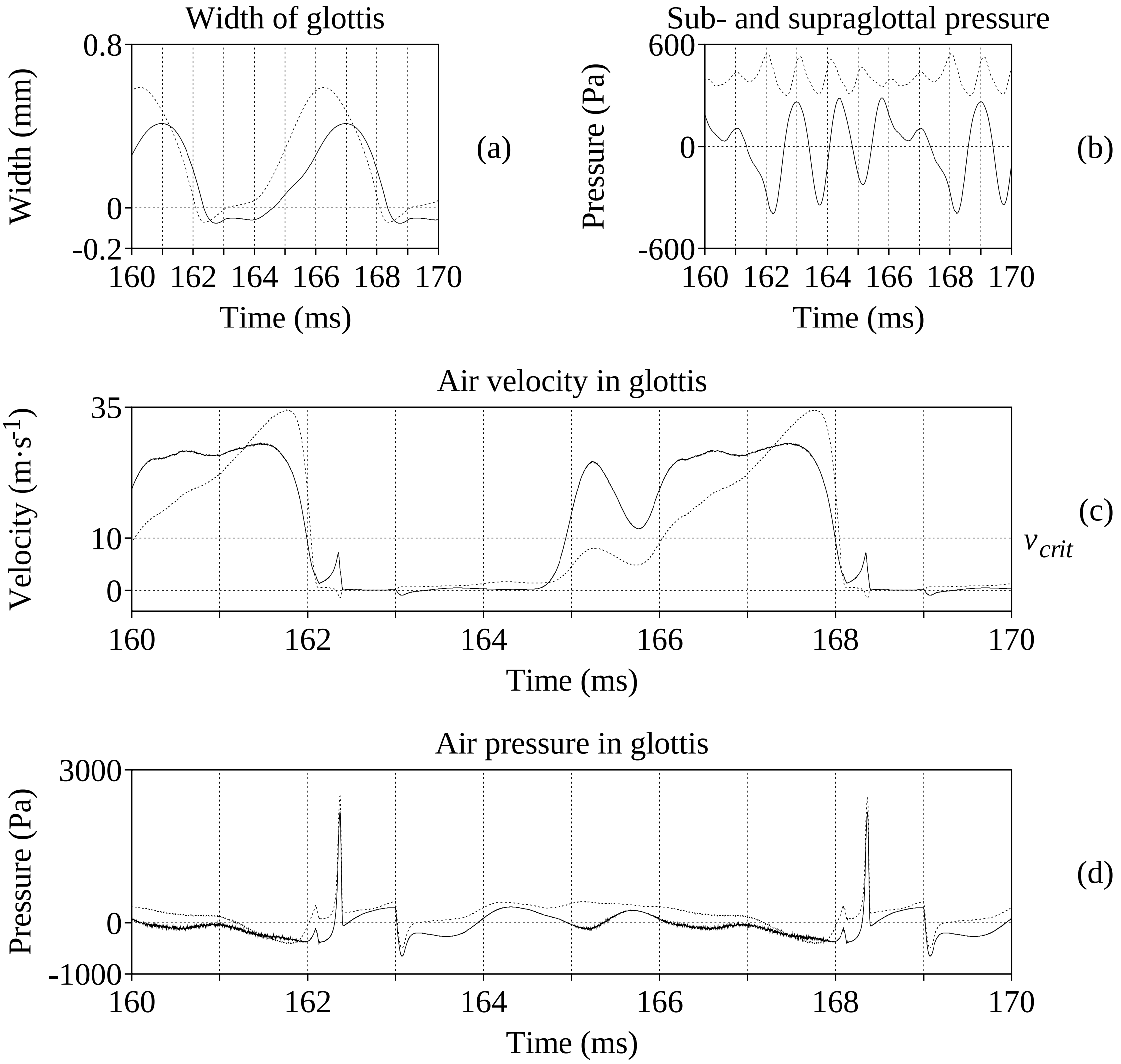


Figure 4. Events in the glottis during phonation. The dotted curves refer to the lower part of the glottis or to the subglottal pressure, and the solid curves refer to the upper part of the glottis or to the supraglottal pressure.

EVENTS IN THE GLOTTIS

Because the glottis is described in the same way as the vocal tract, we get a detailed view of the air pressures and particle velocities in the glottis. Phonation (fig. 4) automatically results when the lungs contract while the vocal cords are adducted and the supralaryngeal pathway is unobstructed.

Figure 4a shows that the glottis is closed about half the time.

Fig. 4b clearly shows the subglottal and supraglottal first formants. They are more strongly damped when the glottis is open than when it is closed.

Fig. 4c shows that between 165 and 166 ms, the air velocity in the glottis is a direct result of the difference between the formant pressures. We also note in this figure the velocity drop when the glottis closes at 168.1 ms, and the sudden velocity peak arising at 168.3 ms between the closing upper parts of the vocal cords when the last amount of air is forced into the pharynx while the lower glottis is already closed.

If the velocity is above 10 m/s, noise is generated, as we see in the upper glottis between 166 and 168 ms (fig. 4d).

Also in fig. 4d, we see the reason why the upper parts of the vocal cords hesitate to open between 163 and 164 ms (fig. 4a), thus causing the long closure interval: as the lower glottis opens, the air is rarefied there and the pressure drops to negative values; this sucks the upper parts of the vocal cords together. In the two pressures of fig. 4d, we also see the formants, and positive pressure peaks when the vocal cords collide.

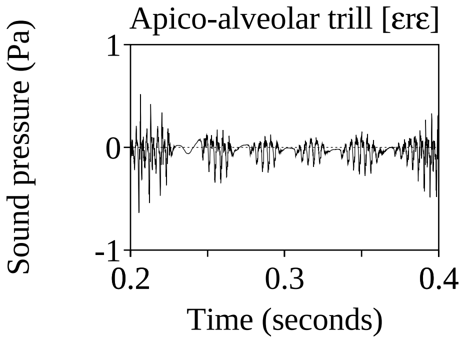


Figure 5. Vibration outside the glottis.

TRILLS

Because the vocal tract is described in the same way as the glottis, our model speaker can generate apical, labial, and uvular trills as easily as vibrating vocal cords. This is shown in fig. 5.

FALSETTO

Our model speaker's voice enters the falsetto register when, other things staying equal, the lung pressure becomes very low. Fig. 6 shows that the lower parts of the vocal cords do not close. The register break occurred when the pressure had fallen to 250 Pa during sustained phonation. Incidentally, our speaker managed to phonate for 20 seconds starting from only 400 Pa without adjusting her lung width, which is much longer than our *male* model speaker can do (11 seconds) and also longer than in reality. This sugggests that a realistic modelling of the amount of air that leaves the lungs during phonation, can only be achieved if we model a substantial leak parallel to the glottis, especially for female speakers. Our model can easily handle this if we add two three-way boundaries (fig. 2), but this was not used for the present paper.

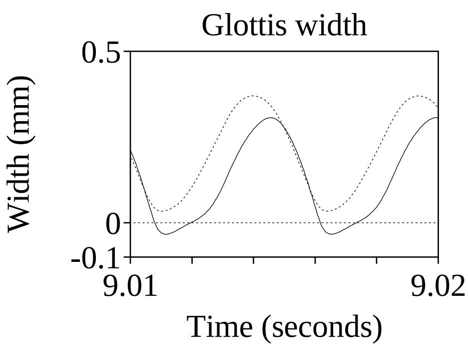


Figure 6. Falsetto at a mean transglottal pressure of 200 Pa. The upper part of the glottis (solid curve) closes, but the lower part (dotted curve) does not.

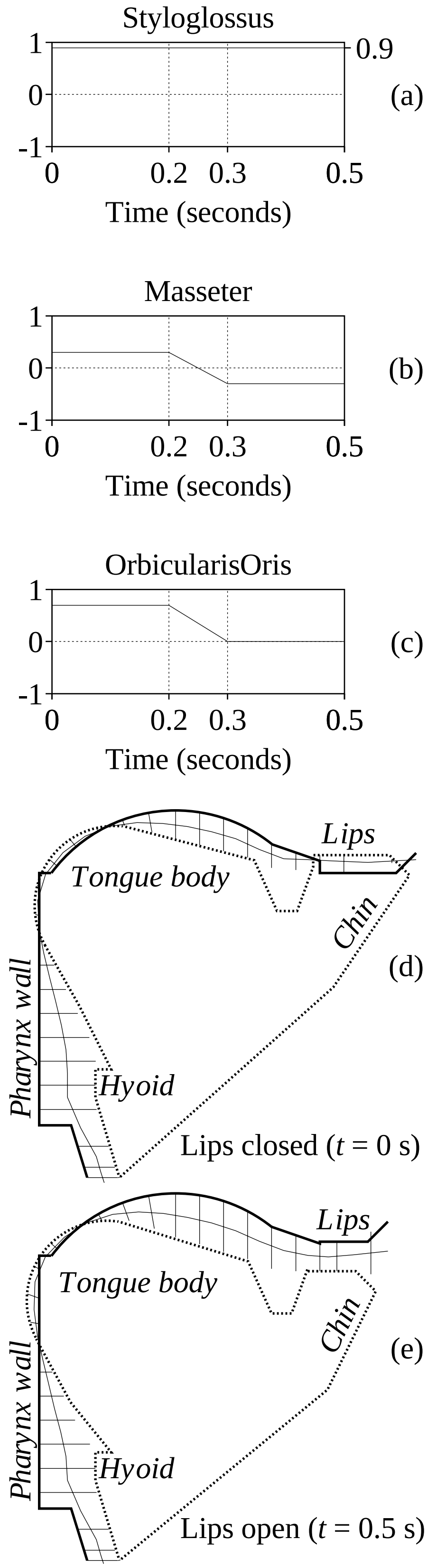


Figure 7. Input for a bilabial click: muscle activities and target shape. Styloglossus pulls the tongue back up. "Masseter" stands for all the muscles that close or open the jaw. Orbicularis oris rounds and protrudes the lips.

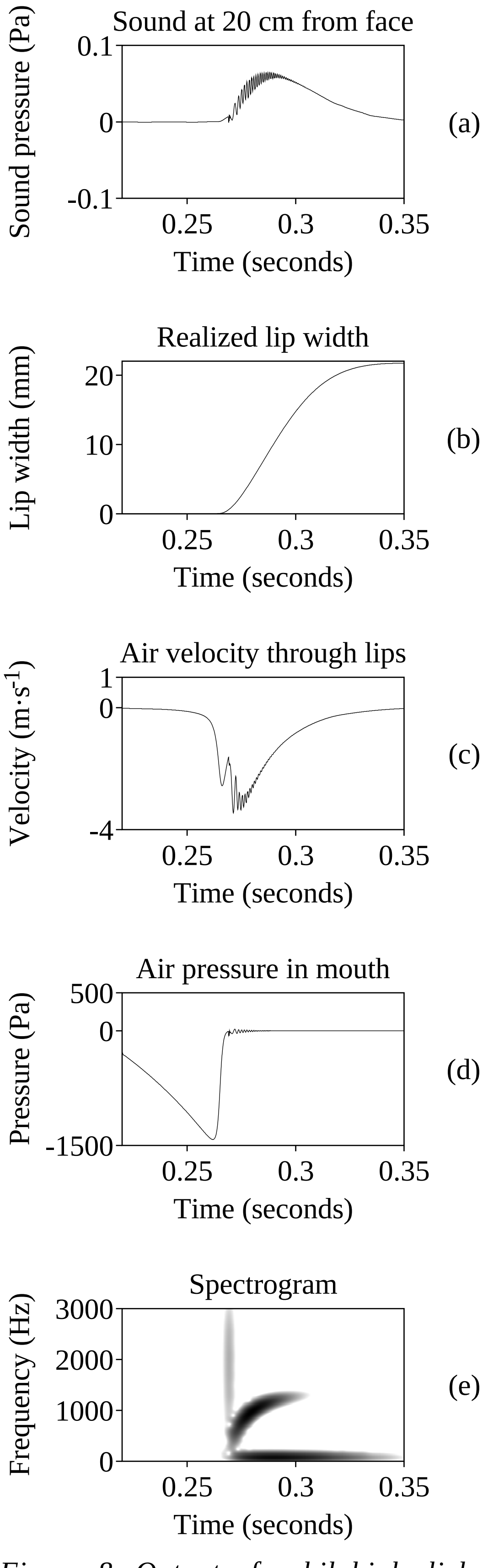


Figure 8. Output of a bilabial click: sound, movements, and aerodynamics.

VOICING IN OBSTRUENTS

Our model speaker can make an [aba]-[apa] contrast by varying her oral wall stiffness only [1]. This suggests that for implementing voicing contrasts languages can use supralaryngeal myo-elastic features, apart from more drastic measures like aspiration or constriction.

A BILABIAL CLICK

Because of the realistic modelling of sucking effects, our model speaker can produce click consonants. Figure 7 shows how our model speaker lowers her jaw and unrounds her lips while maintaining a velar closure. The lungs and glottis are not involved.

Figure 8 shows the acoustic, aerodynamic, and myoelastic results. Jaw lowering starts at 0.20 seconds (not in the figures), which causes the air pressure in the mouth to fall to -1500 Pa relative to the atmospheric pressure (fig. 8d). At 0.27 seconds, the lowering of the jaw causes the lips to separate (fig. 8b), which causes air to flow from outside into the mouth through the lips (fig. 8c). This air flow, which reaches a velocity of 3 m/s, quickly restores the air pressure to the atmospheric pressure (fig. 8d). The resulting sound (fig. 8a) has a tiny burst, which is seen as a vertical band in the spectrogram (fig. 8e; we used a Gaussian window with a -22 dB length of 10 ms, 30 dB dynamic range, and 6 dB/octave pre-emphasis). After the burst, the sound (fig. 8a) shows, superposed on the DC flow, a sine wave with a frequency that rises from 300 Hz to 1200 Hz, which is also reflected in the velocity (fig. 8c) and more clearly seen in the spectrogram (fig. 8e). This *formant transition* is what we would expect for an opening gesture of the lips from an [u]-like position to the [a]-like position of fig. 8b, where the lips end up 22 mm apart.

The auditory impression from the sound (fig. 8a) is correct: the seven listeners that were asked to identify this sound, spontaneously reproduced a non-affricated bilabial click.

REFERENCES

[1] Boersma, P. (1993), “An articulatory synthesizer for the simulation of consonants”, *Proceedings Eurospeech ’93*, pp. 1907-1910.