Mechanical Models of Fricatives Based on MRI-derived Vocal Tract Shapes

Christine H. Shadle[1], Maria Berezina[1,2], Michael Proctor[1], Khalil Iskarous[1]
[1]Haskins Laboratories, 300 George St., New Haven, CT 06511 USA
[2]MIT-SHBT Program, MIT, Cambridge, MA 02139 USA
E-mail: shadle@haskins.yale.edu

Abstract

Articulatory shapes observed in MRI data of fricative production have been used in mechanical models to study the acoustic effects of tongue grooving, upper and lower teeth, size of the sublingual cavity, and position of the tongue relative to a labiodental constriction. The tongue groove shape and extent affects spectral shape and amplitude, particularly at higher frequencies. Inclusion of the teeth increases noise generation, as expected; the amount depends on the tongue shape. The rate of increase of amplitude with flowrate also varied across teeth and tongue models, indicating either differences in the source types or in the efficiency of exciting tract modes. These experiments should help us identify the contribution of each element of a more realistic vocal tract geometry, and allow us to better understand acoustic features of most use in any comparison of fricatives.

1 Introduction

Mechanical models have often been used to study fricative consonants because turbulence noise resists an analytical solution. Early models were used to determine the source characteristics for use in a transmission-line analogue of the vocal tract [2]. Later model studies showed that while a source could in some cases be separated from the ”filter” and measured directly, this was not possible for all fricative-producing configurations; further, the shape of the tract downstream of the constriction was crucial in determining source characteristics, so that a computer model based on area functions alone could never adequately generate noise sources [9, 8].

Adachi & Honda [1] made mechanical models with a more accurate vocal tract shape by using stereolithography to recreate the three-dimensional vocal tract shapes obtained from magnetic resonance imaging (MRI); they then conducted numerical simulations of airflow through the most anterior 4 cm. However, despite the extremely close match of vocal tract shape to mechanical and computer model shapes, the acoustic outputs did not match so closely, and it was unclear what caused the differences. The nature of the model made it somewhat inflexible experimentally. We know that we need more than an area function specification for fricatives, but it is not clear which aspects of the complex shape of the vocal tract are needed.

In our study, we have also started with MRI data, but have constructed mechanical models differently in order to ensure maximum flexibility. We used MRI data of sustained fricatives from five subjects (three female, two male speakers of American English) to identify shape elements worth exploring via mechanical models. Details of the post-processing of the images and construction of the three-dimensional vocal tract volumes are given in Proctor et al. [6]. High-quality acoustic recordings were made of each subject within a day of the MRI session.

Examination of the data revealed greater than expected articulatory variability between subjects, and corresponding acoustic variability which is discussed in detail elsewhere [5, 7]. Aspects of production involving significant articulatory variation that seemed likely to affect noise generation significantly were singled out for exploration via mechanical models. These include differences in the tongue groove, angle and position of the upper and lower teeth with respect to the exit of the constriction, and the shape and volume of the sublingual cavity. Studies of the effect of glottal shape on the flow separation point ([4]) influenced the choice of tongue groove as a parameter. Earlier work identifying obstacles, dentition, and the curve of the palate as potential locations for sound generation [9] informed the use of teeth models, as did Howe & McGowan's study [3] indicating that the channel between upper and lower teeth may serve to increase the turbulence noise that is propagated to the far-field. The dimensions of the sublingual cavity have been shown to affect the far-field sound significantly [9], and the variation in the sublingual cavity found among our subjects for all sibilants identified this as a useful parameter. Finally, a labiodental constriction was made and observed vowel context effects were mimicked by varying the distance between that constriction and the tongue.

2 Method

Vocal tract models were constructed in a cylindrical polycarbonate tube with an inner diameter of 5.02 cm. Models of different articulators were made from machinable wax and sealed into the tube with silicon adhesive. An example is diagrammed in Fig. 1. The inlet to the tract was PVC tubing, 1.8 cm inner diameter (area = 2.55 cm^2), the end of which was enclosed in an adaptor sealed with O-rings against the inner wall of the polycarbonate tube. Immediately downstream of the inlet tube – the "trachea" – was a tapered orifice, designed to simulate the glottis, formed by a 1 cm thick ring of wax: its minimum diameter was 1.0 cm (area = 0.79 cm^2), with tapered inlet and outlet. The adaptor and glottis could be positioned freely within the tract-tube; in general, they were positioned to give a total tract length from glottis to end of the tube of 19 cm. This length and the inner diameter of the tract tube were chosen to match dimensions of subject M1, the subject with the largest vocal tract.

Airflow was supplied by a Gast air compressor, regulated to a constant pressure, which was then further regulated to supply a controlled volume velocity, monitored by a rotameter. A laminar flow element (LFE; Meriam 50MJ10-9) with a Honeywell pressure sensor was used to measure the instantaneous flow velocity just upstream of the model, and both LFE and model were inside an anechoic chamber. A baffle was created by inserting the end of the

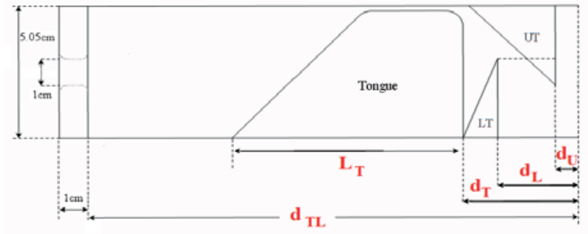

Figure 1: *Diagram of the model with tongue T_{FF}, upper and lower teeth.*

tract into a hole cut out of a plexiglass square (60 by 60 cm). Far-field sound pressure was measured using a Bruel & Kjaer 4190 microphone (with preamplifier B&K 2669 powered by B&K Nexus conditioning amplifier), and the signals from both microphone and the LFE sensor were acquired by a National Instruments data acquisition system at a sampling rate of 44 kHz.

The parameters being manipulated – tongue groove shape, angle and position of teeth, sublingual cavity shape, distance between the tongue and a labiodental constriction – resulted in a number of tongue and teeth models made from cylinders of machinable wax. Tongue models had either a flat constriction, with a flat side in the constriction region giving a constriction cross-sectional shape of a sector of a circle, or a grooved constriction, with a triangular-shaped constriction. The inclined inlet to the constriction was likewise either flat or grooved. In all cases the area of the constriction was designed to be the same, A_C =0.1 cm^2.

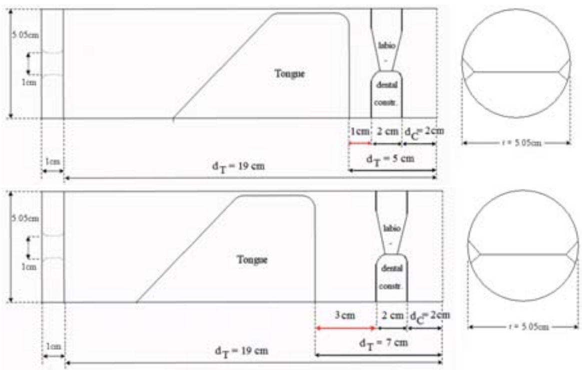

Figure 2: *Two configurations of the labiodental constriction model: tongue anterior (top), posterior (bottom).*

Models of the teeth were machined to have an inclined face upstream and a vertical face (perpendicular to the tract axis) downstream. The upper-teeth model had a 45 degree incline; the lower-teeth model, 60 degree. The lengths were chosen to allow overlap when both were present: upper teeth were 3.0 cm along the tract axis by 3.0 cm across; lower teeth, 1 cm along by 3.0 cm across.

The labiodental model (Fig. 2) was filed by hand to have a thicker lower 'lip' section and an upper section that tapered to meet it. Two constrictions, each of area 0.05 cm^2, were made at the sides.

3 Results

When the different tongue models were placed at the end of the tract (at $d = 0$ cm), the type of constriction (grooved or flat) made the most difference. For the flat constriction (T_{FF} and T_{GF}, shown in Fig. 3), the spectra showed a broad peak typical of a free jet, with low amplitude at low frequencies, a 20 dB rise to a maximum at approximately 5 kHz, and a gradual drop from there to 20 kHz. Grooving on the inlet resulted in a slightly higher amplitude at high frequencies. Contrasting with this pattern, spectra for the grooved constriction (T_{FG} and T_{GG}, not shown) were nearly flat from 4 to 20 kHz. At the lowest flowrate, some whistle-like peaks were observed for the flat inlet-grooved constriction case (T_{FG}).

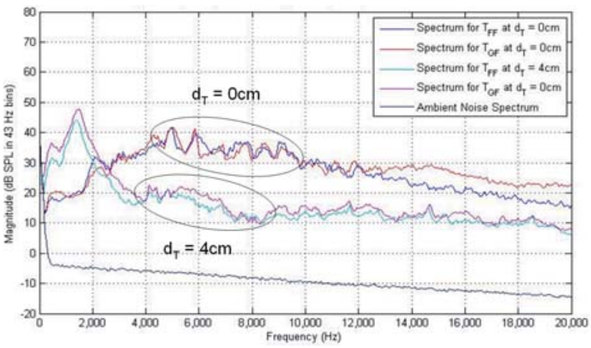

Figure 3: *Spectra comparing tongues with flat constriction, and grooved vs. flat inlet, for 0 and 4 cm front cavity length. Flowrate is 40 l/min.*

When the tongues were placed at $d = 4$ cm, in a more /ʃ/-like position, the differences in amplitude at high frequencies noticeable at $d = 0$ cm tend to disappear (see Fig. 3). In both cases a prominent peak appears at approximately 1.5 kHz, which is the first front-cavity resonance.

Addition of the models of the teeth caused other changes, as seen in Fig. 4. A trough noticeable at 4 kHz in the top graph is likely due to sound generation at the end of the tube; it does not appear when the upper teeth are present, deflecting the jet downwards from the constriction. The front cavity resonance shifts a bit lower in frequency when both teeth models are present (bottom graph), but more important, with teeth the amplitude overall is higher, and the amplitude of the main resonance increases more as flowrate increases, suggesting a different type of source or a more efficient excitation of the tract resonances when teeth are present.

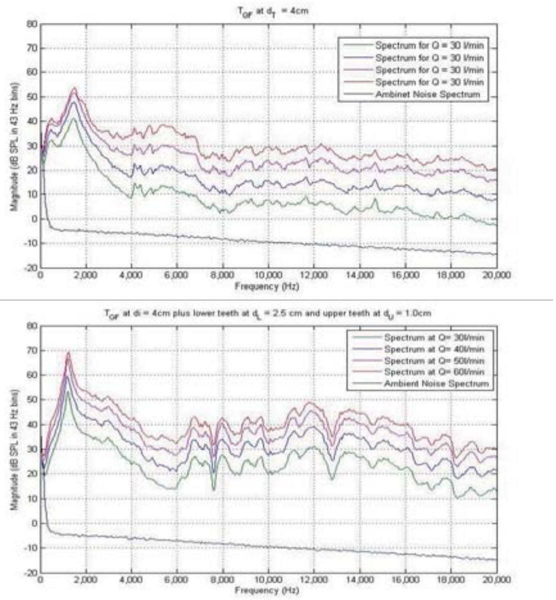

Figure 4: *Spectra of T_{GF} at $d = 4$ cm, flowrates 30, 40, 50, 60 l/min. Top: tongue model only. Bottom: tongue, upper and lower teeth models.*

Midsagittal outlines of subject M2 uttering [f] in three vowel contexts are shown in Fig. 5; this subject showed the greatest effect of vowel context for all fricatives. The model used to study the effect of the different tongue placement in [ufu] and [ifi] is shown in Fig. 2. The spectra resulting at two flowrates are shown in Fig. 6. For frequencies above 4 kHz, the posterior tongue placement resulted in more noise generation, with amplitude difference of 5 - 8 dB. Using T_{GF} instead of T_{FF} resulted in lower amplitudes at all frequencies and a bigger amplitude differential between the anterior and posterior tongue positions. While these models differ from the subject in that the tongue constriction has a fricative-like, rather than vowel-like area, these results show that the configuration posterior to the labiodental constriction can affect the far-field sound significantly.

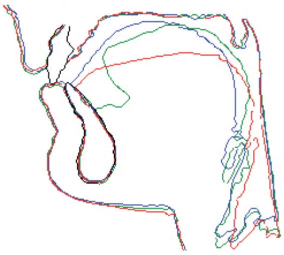

Figure 5: *Midsagittal outlines of [f] for subject M2. Vowel context is red for [afa], green [ufu], blue [afa].*

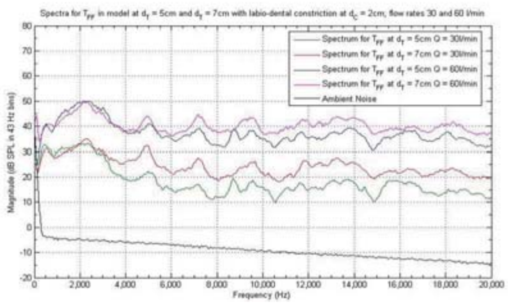

Figure 6: *Spectra of the labiodental constriction with tongue T_{FF} at flowrate 30 and 60 l/min, with anterior and posterior tongue positions.*

The effect of the size of the sublingual cavity was also studied since this varied widely across subjects. Surprisingly, the frequency of the main resonance was not affected. There were some variations in amplitude in the 2 – 5 kHz range.

4 Conclusion

Articulatory shapes observed in MRI data of fricative production have been used in mechanical models to study the acoustic effects of tongue grooving, upper and lower teeth, size of the sublingual cavity, and position of the tongue relative to a labiodental constriction. The tongue groove shape and extent affects spectral shape and amplitude, particularly at higher frequencies. Some configurations show unstable (whistle-like) behavior at lower flowrates. Inclusion of the teeth increases noise generation, as expected; the amount depends on the tongue shape. The lowest-frequency peaks in the spectrum can be predicted from a plane-wave acoustic model; not as easy to predict were amplitude differences, and differences in the rate of increase of amplitude with flowrate, indicating either differences in the source types or the efficiency of exciting tract modes in different configurations. The position of the tongue model, and whether or not its inlet was grooved, affected the spectral amplitude above 4 kHz for the labiodental model.

Acknowledgments: Research supported by NIH NIDCD RO1 DC006705 to the first author and an NIH SHBT Training Grant to MIT.

References

[1] Adachi, S. & Honda, K. CFD approach to fricative sound sources. *Proc. 6th Int'l Sem. on Speech Prod.*, Sydney, 7-10 Dec 2003.

[2] Heinz, J.M. Fricative consonants. *M.I.T. Res. Lab. of Elect. Quart. Rpt.*, Oct-Dec., 5-7, 1956.

[3] Howe, M.S. & McGowan, R.S. Aeroacoustics of [s]. *Proc. Royal Soc.* 461, 1005-1028, 2005.

[4] Pelorson, X., Hirschberg, A., van Hassel, R.R., & Wijnands, A.P.J. Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. *J. Acoust. Soc. Am.* 96, 3416-31, 1994.

[5] Proctor, M., Shadle, C.H. & Iskarous, K. An MRI study of vocalic context effects and lip rounding in the production of English sibilants. *Proc. AICSST*, Auckland, 6-8 December 2006.

[6] Proctor, M., Shadle, C.H. & Iskarous, K. A method of co-registering multiple MR-Imaged vocal tract volumes for fricatives. *Proc. 2nd Joint ASA-EAA Conference*, Paris, 29 June - 4 July 2008.

[7] Shadle, C.H., Proctor, M., and Iskarous, K. An MRI study of the effect of vowel context on English fricatives. *Proc. 2nd Joint ASA-EAA Conference*, Paris, 29 June - 4 July 2008.

[8] Shadle, C.H. The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics* 19:3/4, 409-424, 1991.

[9] Shadle, C.H. Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling* (187-209), Dordrecht: Kluwer, 1990.