

June 29-July 4, 2008

www.acoustics08-paris.org

euronoise

An MRI Study of the Effect of Vowel Context on English Fricatives

C. Shadle, M. I Proctor and K. Iskarous

Haskins Laboratories, 300 George St, New Haven, CT 06511, USA
proctor@haskins.yale.edu

To gain a better understanding of the long-observed effects of vocalic context, the articulation of fricatives was investigated using Magnetic Resonance Imaging. Five speakers of American English were imaged while producing eight fricatives in the contexts /i-a-u-ə/. Sagittal, axial and oblique-coronal volumes were acquired for each vowel-fricative combination. Acoustic recordings were made during scans and separately in an anechoic chamber. Vocal tract models were generated by aligning and superimposing all three stack orientations. The models reveal that a variety of articulatory strategies are employed in the production of English fricatives, and that vocalic context is significant.

For some subjects, tongue shape differs little with vowel context; other subjects show highly varied tongue shape differences but little difference in lip rounding. Two subjects show significant variation with vowel context for every fricative, including [θ]. Two show very little difference, even for [f]. The sublingual cavity in [ʃ] is extensive for two subjects, but only slightly bigger than in [s] for two other subjects. Tongue grooving in and behind the constriction also varies, and is especially evident in the oblique scans. Overall, more variation and effect of vowel context was observed than expected from the literature.

1 Introduction

In modeling fricatives, it has long been known that the size and shape of the constriction, and the three-dimensional shape of the vocal tract downstream of the constriction, are crucial [9]. Magnetic Resonance Imaging (MRI) is therefore in some ways well-suited to imaging the vocal tract during fricative production, because it is safe, visualizes soft tissue clearly, and generates a stack of image slices in a user-chosen plane, allowing construction of a true three-dimensional volume. There are drawbacks, however: imaging time is long for each stack, scans are expensive, and the teeth appear no different from air.

A 1995 MRI-based study of fricative articulation [6] used four subjects and three image-plane orientations, sagittal, axial and coronal, for each fricative. The fine spatial resolution chosen resulted in scan times for each stack of 1.5-2.5 minutes. As a result, each fricative was imaged in a single neutral vowel context. However, many studies have shown that the acoustic properties of fricatives are affected by vowel context (e.g. [4, 10]), and a 1996 MRI study has shown that vowel context effects were observable even during the long holds necessary for MRI [11]. Further, while some studies of fricatives using acoustic data only have demonstrated much variability across subjects, we would like to know if such variability can be explained by differences in the articulatory configurations.

We obtained MR images of fricatives in four vowel contexts for five subjects, and corresponding acoustic recordings. In this paper we focus mainly on articulatory effects of vowel context. First, we review articulatory findings from the literature.

2 Background

Many studies exist of the articulatory properties of English fricatives. Subtelný et al. [13] studied [s,z] in 10 subjects. Although their use of cine X-rays meant that the corpus had to be short, they chose a sentence in which [s] occurred in 4 vowel contexts and [z] in one, and also included a sustained [s]. They found a high degree of precision in the constriction location and length across subject, but more variation in tongue body and lip parameters both across subject and with vowel context. The phonetic context affected mainly pharyngeal

tongue shape and lip opening; the tongue root was more advanced for the [z]. Their results were consistent with the constriction being formed by tongue tip or blade, but reported more consistency across subjects than expected from the literature.

Catford [1] described [θ, s, ʃ] using direct palatography as well as radiographic data. The channel width was wide for [θ], narrow for [sh] and narrowest for [s]. He noted that while [θ] is always apical, [s] can be apical or laminal, and discussed how this difference affects the presence of a sublingual cavity and the shape of the tongue behind the constriction. An apical [s], if alveolar or further back, generates a small sublingual cavity; if the [s] is postalveolar, a laminal articulation will have a dome-like palatalized tongue, whereas an apical articulation will very likely have a sulcus behind the constriction, and doming towards the velum.

Catford's description conflicts with data presented by Perkell et al. [7]. They measured the strength of contact of the tongue on lower teeth in various tokens of [s, ʃ], including combinations across word boundaries. For their four subjects, they found that [s] always had strong contact, whereas [ʃ] ranged from partial to no contact, depending on the subject. They argued that any amount of sublingual cavity changed the acoustic properties significantly, producing resonant frequencies characteristic of [ʃ]. Though their subjects differed somewhat from each other, it would seem that none of them was similar to the subjects Catford described as having apical, alveolar [s] productions with small sublingual cavities.

Fletcher and Newman [2] used EPG data from three subjects to study tradeoffs in constriction location and groove width in [s-ʃ] production. They noted that groove width for the two subjects was very similar, even across six vowel contexts; [ʃ] always had a wider groove. However, place relative to the incisors varied. When the subjects were asked to vary place and groove width deliberately, no one parameter affected the perceived result consistently across subjects, indicating that these are joint articulatory features for [s] and [ʃ].

Narayanan et al.'s [6] MRI study used four subjects, and [ə] context only. They reported that the alveolars had a concave tongue-body shape, while the postalveolars had a convex shape. The amount of concavity was greater for apical alveolars; whether the articulation was apical or laminal was subject-dependent. The

nonstrident fricatives were more variable in the pharyngeal region than were the stridents.

Shadle et al. [11] used MRI to study vowel context of one American English subject. In spite of the duration of the scan (approximately 2.5 minutes per image stack), an effect of vowel context was observed. Tongue position varied most for [f] and least for [ʃ].

Stone and Lundberg [12] used ultrasound and EPG data of one subject to identify shape factors describing the tongue for all vowels and nearly all consonants of American English. For their subject, /s-f-θ/ tongue shapes were defined. [ʃ] has front raising; [θ, s] have complete groove through to tongue tip. In [ʃ], the groove is posterior of the level or arched coronal part of tongue. [θ, s] had ‘essentially the same shapes’.

Ladefoged and Maddieson [5] summarize many studies. Among their conclusions for American English, they note that the teeth are close together for [s, ʃ]. The tongue position can vary for [θ], protruded between the teeth or resting just behind the upper teeth. Various studies indicate that the tongue tip can be up or down for [s]. Further, they describe a “deep pit” that may occur in [s] behind the groove, whereas the tongue is domed behind the constriction for [ʃ]. The [ʃ] constriction is wider and further back than in [s]; the place of constriction for [s] can vary depending on a number of factors such as the rugae, shape of the palate, and shape of the teeth.

To summarize the expectations, for each fricative place there are different sets of primary and secondary articulators. The secondary articulators are more apt to coarticulate with surrounding vowels. Thus in [f] the tongue can be expected to show the effect of vowel context the most. In both [θ] and [s] the tongue tip or blade is the primary articulator, but the tongue dorsum and lips may coarticulate. In [ʃ] the lips generally protrude or round in American English, and so may show vowel effects less. On the other hand, [s] has been called the most resistant to coarticulation generally, though that should apply to tongue position and tongue grooving, not lip position. Hardcastle [3] describes the [s] as requiring a complex articulation involving delicate tradeoffs, but notes that [f] possibly has as many tradeoffs involving the labial muscles.

There is consensus that /s/ can be apical or laminal, has a deep and narrow groove. The shape of the tongue posterior to the constriction can involve a depression or sulcus, with doming posterior to that, but this shape depends on where the /s/ is articulated and whether apical or laminal.

It is expected that [ʃ] is always laminal, and always has a sublingual cavity; /s/ may or may not. [ʃ] also has a groove, though not as narrow as for [s], and it does not continue through the constriction. The groove characteristics may depend on the place for [s, ʃ] in a given subject. [θ] likewise has a groove that continues through the constriction, but is wide.

3 Method

Magnetic Resonance images, acoustic recordings and dental impressions were obtained for five subjects (2 men,

M1, M2; 3 women, W1-W3), who are native speakers of American English. All were linguistics students so that the corpus could be presented in IPA. The corpus consisted of each of the English oral fricatives [f-v, θ, ð, s,z, ʃ, ʒ], each elicited in the four vowel contexts [i-a-u-ə]. Subjects were instructed to concentrate on maintaining a consistent vocalic context throughout, to sustain even frication throughout the production, and to ensure that they pronounced the vowel at the end of the fricative after the scan had stopped.

A Siemens Sonata 1.5T MRI Scanner was used to image the subjects’ vocal tracts while they produced all fricative tokens over the course of two or three 90-minute sessions. Subjects lay supine in the scanner, sustaining each fricative in each vocalic context for 36 seconds. Prompts were presented in IPA, projected onto a screen which could be read by the subject from within the scanner bore. Some subjects interrupted the sustained frication in order to take one breath; they were instructed to do so with a minimum of oral movement to reduce image blur.

A 2D True-FISP scan sequence ($T_r = 200$ ms, $T_e = 3.3$ ms, Flip Angle = 70°) was chosen as the best compromise between image resolution and scan time. Each token was repeated three times, once for each image orientation: sagittally (from ear to ear), axially (upper trachea to inferior nasal cavity), and obliquely, using axial planes tilted to 45° , providing cross-sectional imaging of the tract in the alveolar region (lips to velum). In each orientation, slices of 4 mm thickness were used, spaced at 4.8 mm intervals.

The corpus was split initially into two sessions so that all scans of /f,v,s,z/ were taken in one session, and all of /θ, ð, ʃ, ʒ/ in the other. In some cases a third session was necessary. Subjects were monitored during the scans using an optical microphone adaptive noise cancelling system (Phone-Or and MR-Confon) to ensure that the correct utterance was produced. Any tokens that were incorrectly produced or for which the scans looked blurry due to subject motion were rescanned.

In addition to the recordings made during the scans, subjects were recorded in an anechoic chamber with a day or so of the imaging session. They produced the MRI corpus using the same IPA prompts, and also produced each fricative-vowel combination at shorter durations, either 10 or 5 seconds long, to allow checks to be made on the naturalness of the 36-second-long tokens. They were recorded using a Bruel & Kjaer 4190 condenser microphone with 2669 preamplifier and a Nexus amplifier and filter set to bandpass filter from 20 Hz to 22 kHz. The signal then was sampled at 44 kHz by a National Instruments-based data acquisition system. The spectra shown in this paper are periodograms computed from a 5-sec section edited from the middle of each long-hold fricative.

For each subject, image stacks from each orientation for a given fricative-vowel combination were processed to result in a combined three-dimensional representation of the vocal tract shape. The details are given in the companion paper [8].

4 Results

Overall, the vowel context does have an effect on both the articulation and acoustic output of subjects, but the amount and type of effect varies by subject. In general, M1 and W1 showed little effect; W2, some; and M2, W3 quite a bit. Although we analyzed all four vowel contexts, in the results below we discuss only the cardinal vowels because /ə/ appeared to pattern with a different vowel for each subject. We now consider the vowel context effects in some detail, discussing each fricative place in order.

4.1 Effect of vowel context on [f, v]

Figure 1 shows data for two subjects, M2 and W2. The outlines of the midsagittal contours for [a,i,u] contexts for [f] are superimposed for each subject. M2 shows the predicted result, with the tongue position that of the surrounding vowel in each case. W2 instead shows a very similar tongue position for all three contexts, a high back position approaching that of [u]. The other subjects likewise used a single tongue position for all [f] tokens, but the position differed by subject: neutral for M1, [u]-like for W1, and [i]-like with a large sublingual cavity extending under the tongue tip for W3.

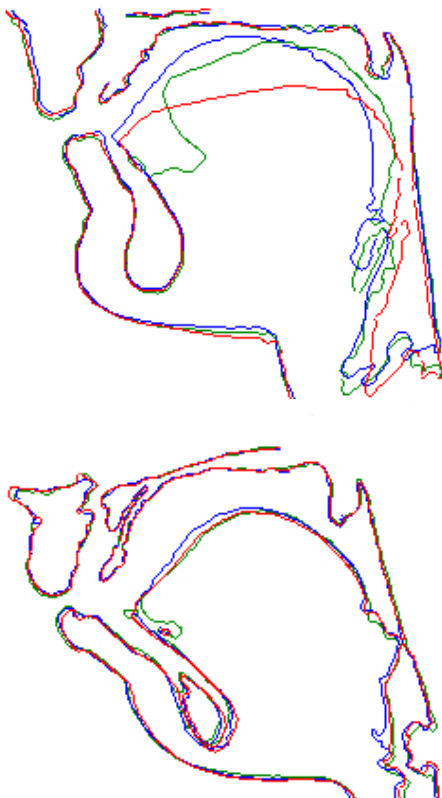


Figure 1: Midsagittal outlines for subjects M2 (top) and W2 (bottom), for [f] in contexts [afa] (red), [ifi] (blue), [ufu] (green).

For [v], vocal tract configurations were similar in most cases to that subject's corresponding voiceless token, except that the tongue was always advanced for [v]. For M1, the tongue was slightly lower for [iv:i] than [av:a, uv:u]. For W3, the size of the sublingual cavity

varied by vowel for [v], with no cavity visible under the tongue tip for [iv:i].

4.2 Effect of vowel context on [θ, ð]

Figure 2 shows midsagittal outlines for [iθ:i, aθ:a] for subjects M1 and M2. M1 shows almost no difference by vowel, with a convex tongue surface; M2 shows a noticeably higher tongue position for [iθ:i] than [aθ:a], with a corresponding concavity behind the tongue apex in the former case. Subject W1 had similar results to M2. Subjects W2 and W3 showed essentially no differences by vowel, but both showed a similar concavity behind the tongue apex.

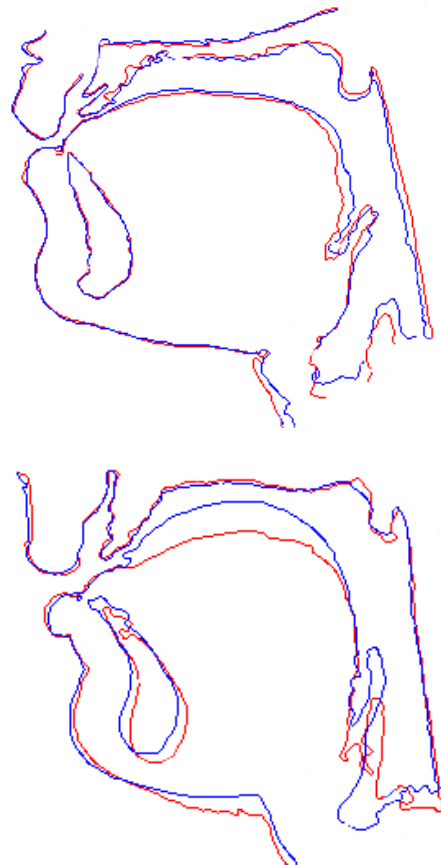


Figure 2: Midsagittal outlines for subjects M1 (top) and M2 (bottom) for [θ] in vowel contexts [a] (red), [i] (blue).

Results for the voiced cases were again similar to the voiceless, except for the more advanced tongue root. Subject M2 still showed a significant difference by vowel context, though not quite as much as for [θ].

4.3 Effect of vowel context on [s,z]

For [s], subject M2 again shows the most variation with vowel context, but none in the region of the constriction, as expected. Instead the concavity in the tongue behind the constriction is deepest for [us:u], shallowest for [is:i], and the lower lip is more protruded for [us:u]. All subjects show some degree of concavity behind the constriction, even M1, whose articulation could be described as laminal.

Figure 3 shows the tongue groove for [s, ʃ] in [i] context for W3. As is the case for most subjects, the groove is deeper for [s]. For W2, however, the groove extends into the constriction for [is:i] but stops short of it for [as:a]. The other subjects show a groove that deepens anteriorly and is not affected strongly by vowel context.

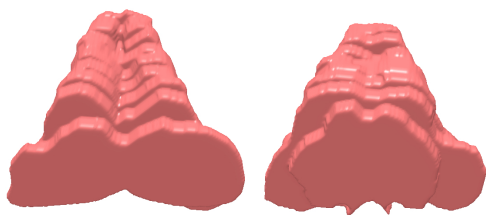


Figure 3: Subject W3; the oblique scans were used to create a superior view of the tongue groove. Tongue tip is at the top. [is:i] (left) and [iʃ:i] (right).

In the voiced cases, tongue root is again advanced relative to the corresponding voiceless cases; the effect of vowel context is otherwise similar within each subject.

4.4 Effect of vowel context on [ʃ, ʒ]

For M1 and W1, there was no difference by vowel context apparent from the midsagittal slices, and for W2, there was only a slight difference in the vicinity of the constriction for [iʃ:i]. However, the oblique slices for W2 revealed a pattern much like that observed in [s]: the groove was more pronounced, and continued through the constriction, in [iʃ:i] compared to [aʃ:a]. For this subject, the difference in the groove from vowel context is more significant than that from [s] vs. [ʃ].

Subjects W3 and M2 showed significant effects of the vowel context, as shown in Fig. 4. For W3, the size of the sublingual cavity and the depth of the concavity behind the constriction varied; there was no groove visible for [iʃ:i]. For M2, the sublingual cavity also varied, but the tongue was only inflected for [aʃ:a], thus making the constriction in that case the shortest of the three.

In the voiced case, most subjects showed the same patterns as in the voiceless case, but with a more advanced tongue root. For M2, however, the vowel context had some effect on the tongue shape for [ʒ], but much less than for [ʃ].

4.5 Other findings

Detailed comparison of acoustic spectra to the articulatory configurations is outside the scope of this paper, but we do make a few comments here. The subject with the smallest vocal tract and most anterior tongue position for all fricatives, W2, also showed spectral peaks at the highest frequencies, as one would expect. However, in many cases the acoustic variation (or lack of it) was not easily predictable from the articulatory variation.

As one example of this, Figures 4, 5, and 6 show midsagittal outlines and the corresponding spectra for subject W3's [s, ʃ]. Briefly, she shows the maximum articulatory variation by vowel context in [ʃ], but the maximum acoustic variation by vowel context in [s].

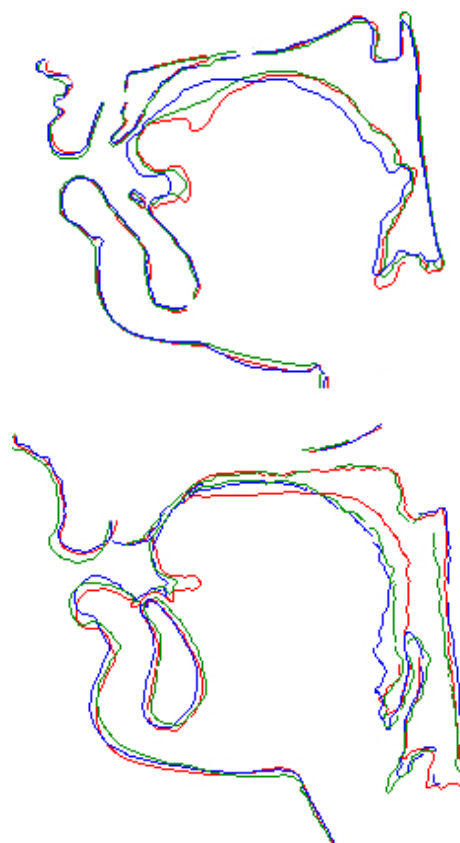


Figure 4: Subjects W3 (top) and M2 (bottom); midsagittal outlines of [ʃ] in vowel contexts [a] (red), [i] (blue), [u] (green).

5 Discussion and Conclusion

In this paper we studied the effect of vowel context on fricatives, using MRI images and accompanying acoustic recordings for five speakers of American English. There is lots of variation in effect of vowel context on fricatives.

Fricatives in the schwa vowel context are the most variable across subject because the subjects do not use the same “neutral” vowel in each case. This may explain some of the variability Narayanan et al. [6] observed.

As Hardcastle stated [3], [f] does allow the most freedom for the tongue to take on the position of the surrounding vowel. However, four of our subjects used a single tongue position for all [f] productions; only one showed the expected variation with vowel context. In [θ], by contrast, where one might expect the least variation with vowel context or across subject since the tongue is pinned at the teeth, two subjects show considerable variation in tongue dorsum position.

As did Subtelný et al. [13], we found that some subjects used a laminal and some an apical articulation for [s]. Subject M1 used a laminal articulation for nearly all fricatives; subject W1 did as well, except for [s]. However, the depth of the depression behind the constriction did not seem to depend on the degree of apicality but rather on the vowel context.

Subject W3 shows that [ʃ] is not always domed. The use of the oblique scans showed detail on the grooving pattern. Observations by Subtelný et al. and Stone and Lundberg fit some but not all of our subjects.

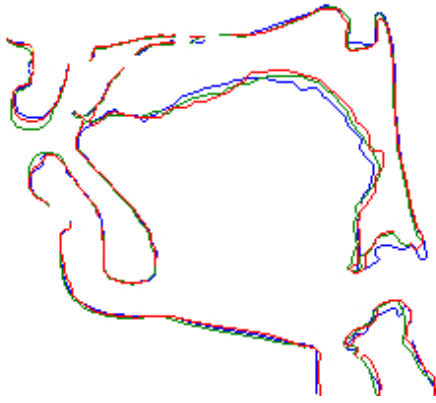


Figure 5: Subject W3, midsagittal outlines of [s] in vowel contexts [a] (red), [i] (blue), and [u] (green).

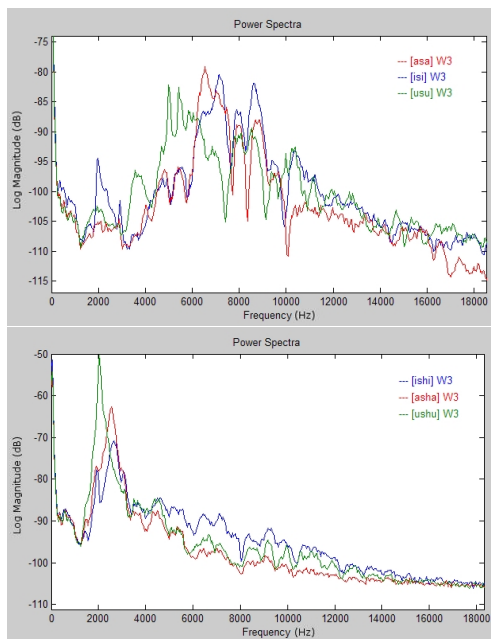


Figure 6: Spectra for W3. Top: [s], bottom: [ʃ] in three vowel contexts. [aF:a] = red, [iF:i] = blue, [uF:u] = green.

While the front cavity was always, for a given subject, bigger for [ʃ] than for [s], the idea that [ʃ] has a sublingual cavity and [s] doesn't is too simplistic. Two subjects (W3, M2) had pronounced sublingual cavities that extended under the tongue tip and varied with vowel context; two subjects (W1, W2) had very anterior tongue positions and so a relatively small front cavity volume. For [ʃ], our results show that constriction length can vary widely within a subject. Constriction exit location seems more constant across vowel context within a subject, which would tend to reduce the acoustic variability; however, that exit location varies across subjects quite considerably for both [ʃ] and [s].

While articulatory and acoustic variability in fricatives has been reported before in the literature, we were somewhat surprised by the number of exceptions we found in our five subjects to generally accepted aspects of fricative production. We can only conclude that such generalizations have been hampered by the difficulties of obtaining articulatory data, which have tended to re-

sult in a small number of subjects, small corpus size, or both.

The relation of articulatory to acoustic variation is not at all straightforward. This will be reported on more in future.

Acknowledgments

The work presented here is part of an ongoing project supported by NIH-NIDCD-RO1-DC006705.

References

- [1] J.C. Catford, *Fundamental Problems in Phonetics*. Bloomington: Indiana University Press (1977).
- [2] S.G. Fletcher, D.G. Newman, “[s] and [sh] as a function of linguapalatal contact place and sibilant groove width,” *JASA* 89, 850-858 (1991).
- [3] W.J. Hardcastle, *Physiology of Speech Production*. London: Academic Press, 1976.
- [4] J.M. Heinz, K.N. Stevens, “On the properties of voiceless fricative consonants,” *JASA* 33, 589-596 (1961).
- [5] P. Ladefoged, I. Maddieson, *The Sounds of the World's Languages*. Cambridge: Blackwell Pub. (1996).
- [6] S. Narayanan, A. Alwan, K. Haker, “An articulatory study of fricative consonants using magnetic resonance imaging,” *JASA* 98:3, 1325-1347 (1995).
- [7] J.S. Perkell, S.E. Boyce, K.N. Stevens, “Articulatory and acoustic correlates of the [s-sh] distinction,” *Speech Communication Papers*, eds. J.J. Wolf and D.H. Klatt, New York: ASA, 109-113 (1979).
- [8] M. Proctor, C. Shadle, K. Iskarous, “A method of co-registering multiple MR-Imaged vocal tract volumes for fricatives,” *Proc. of Joint ASA-EAA Conference*, Paris (2008).
- [9] C.H. Shadle, “The effect of geometry on source mechanisms of fricative consonants,” *J. Phon.* 19, 409-424 (1991).
- [10] C.H. Shadle, S. J. Mair, J.N. Carter, “Acoustic characteristics of the front fricatives [f,v,θ, ð],” *Proc. 4th Speech Prod. Sem.*, Autrans, 193-196 (1996).
- [11] C.H. Shadle, M. Tiede, S. Masaki, Y. Shimada, I. Fujimoto, “An MRI study of the effects of vowel context on fricatives,” *Proc. IOA* 18:9:1, 187-194 (1996).
- [12] M. Stone, A. Lundberg, “Three-dimensional tongue surface shapes of English consonants and vowels,” *JASA* 99, 3728-3737 (1996).
- [13] Joanne D. Subtelny, N. Oya, J.D. Subtelny, “Cineradiographic study of sibilants,” *Folia Phon.* 24, 30-50 (1972).