An MRI study of vocalic context effects and lip rounding in the production of English sibilants

Michael Proctor, Christine Shadle and Khalil Iskarous

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

Abstract

An MRI analysis of [s] and [ʃ] has shown that the articulatory strategies employed in the production of English sibilants vary considerably between individuals and according to vocalic context. Greater variation was observed in the production of [ʃ] than for [s], which was consistently produced with a more anterior constriction and no detectable sublingual cavity. [ʃ] was generally articulated with the tongue blade distributed behind the alveolar ridge; one subject used a more retroflexed tongue posture to create a less distributed constriction in the same region. Lip rounding was evident in the production of [ʃ] by three subjects; for two subjects the difference in lip rounding between [s] and [ʃ] was negligible. Significant acoustic differences corresponding to differences in place of articulation were observed for [s], and for the tokens of [ʃ] which were produced with a pronounced sublingual cavity.

1. Introduction

It has long been observed that the articulatory and acoustic properties of [s] and [ʃ] depend on phonetic context. Changes in the place and area of constriction of [ʃ] can be attributed to vowel context (Shadle, Mair, Carter & Millner, 1995), and vocalic context effects have been shown to have a significant influence on [s], especially where the rounding feature of the vowel extends into neighbouring fricatives (Shadle & Scully, 1995). This effect is of particular interest when considering the way in which [s] and [ʃ] are differentiated, if rounding is a distinctive feature of the post-alveolar fricative.

The interpretation of these effects is complicated by our incomplete understanding of the relative importance of the various articulations involved in the production of the two sibilants. Perkell, Boyce & Stevens (1979) proposed that while both [s] and [ʃ] require the creation of a groove in the tongue blade, the articulation of [s] might involve a "relatively imprecise (anterior) gesture" of the tongue blade, and [ʃ] a "relatively precise positioning of the tongue tip along the front-back dimension". This analysis differs from other theories which consider both sibilants to be articulated by the formation of a similar type of constriction, differing only in location.

A larger question which underpins the study of sibilants is the debate over the essential nature of fricatives in general. Acoustic theories of phonology, such as Stevens (1998), claim that the essential qualities of a fricative are ultimately acoustic and therefore lie in the spectral qualities of the phone. Articulatory theories (e.g. Fowler, Rubin, Remez & Turvey, 1980; Browman & Goldstein, 1989) propose that the fundamental properties of a given fricative will ultimately be gestural, and that acoustic properties serve only to allow the listener to reconstruct the gestures of the speaker.

Sibilants provide an interesting class of sounds with which to test some of the issues central to this debate since they resist simple acoustic or articulatory characterisation. Cohen & Perkell (1986), for example, argue that [ʃ] is characterised by an acoustic signature with a spectral peak around 2.3kHz, which contrasts with the higher frequency peak of [s] (5kHz). However the authors found that [ʃ] is produced by "the sudden creation of a sublingual space", which suggests that these spectral qualities may also be seen as the acoustic signature of a characteristic gesture which differentiates this phone from the alveolar fricative.

Further insights into the articulatory nature of fricatives have recently been facilitated by the use of Magnetic Resonance Imaging (MRI) of the vocal tract. The first MRI study of fricatives (Narayanan, Alwan & Haker, 1995) used only schwa as a vocalic context; however, the effect of vocalic context is apparent even in the sustained fricatives used in MR imaging of the vocal tract (Shadle, Tiede, Masaki, Shimada & Fujimoto, 1996).

The objective of this study is to investigate the articulation of English sibilants, and the influence of neighbouring vowels more thoroughly, by using a full set of vowel contexts, and a wider range of MR imaging and analysis techniques than have previously been employed.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

2. Method

2.1. Subjects

Five native speakers of Standard American English, three women (subjects F1-F3) and two men (subjects M1, M2), were recruited for the experiment. All subjects were aged between 21 and 26 years. None were bilingual or had lived more than 18 months outside of the US, although one subject was fluent in Spanish and all had some basic knowledge of at least one other language. All subjects had normal dentition; none spoke with any speech defects. All subjects were IPA-literate students of linguistics who were paid for their participation and trained in each task beforehand. Non-naïve subjects were deliberately chosen so that stimuli could be presented in IPA, and the subjects could be instructed about the linguistic requirements of each task (e.g. "ensure that the vowel is as high and front as possible"; "try to maintain even frication throughout").

2.2. Corpus Design

The sibilant corpus considered in this paper represents a subset of data collected for a larger study of the full set of English fricatives. Each fricative was elicited in four different vocalic contexts – three maximally distributed cardinal vowels and central schwa – making a total of eight tokens. The corpus was elicited from each subject using the following stimuli: [pisːi], [pasːa], [pusːu], [pəsːə], [piʃːi], [paʃːa], [puʃːu], [pəʃːə]. Subjects were instructed to concentrate on the vocalic context throughout the fricative, and to ensure that they pronounced the vowel at the end of the sustained frication.

Although both sibilants were elicited in the token [pəSːə], the quality of schwa was found to differ too much between subjects to provide a reliable central vocalic context. For some speakers schwa patterned with the low vowel as a kind of lax [ɐ], while for other subjects it was realized as the higher [ɵ]. For this reason only those sibilants produced in the context of one of the three cardinal vowels [i_i], [a_a], [u_u] were considered in the analysis which follows.

2.3. Vocal Tract Imaging

A Siemens Sonata 1.5T MRI Scanner was used to image the vocal tracts of the subjects while they produced both sets of fricatives during a 90 minute session. Subjects lay supine in the scanner and sustained each fricative in each vocalic context for 36 seconds. In most cases the subject 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. Prompts were presented in IPA on a projector screen, which could be read by the subject from within the scanner bore. Presentation of stimuli was timed to coordinate with the scanner sequence to minimise subject movement during scanning, limiting motion blurring.

A 2D True-FISP scan sequence (T_r=200ms, T_e=3.33ms, Flip Angle=70°) was chosen as the best compromise between image resolution and scan time. A sagittal scan sequence acquired 21 slices of 4mm thickness, spaced at 4.8mm intervals, imaging the entire vocal tract from ear to ear. Each utterance was also imaged using an axially-oriented scan sequence (parallel to the subject's bite-plane) with the same parameters. Selected tokens were also imaged using an oblique scan, oriented at 45° to the axial planes, which provided superior cross-sectional imaging of the area of the tract around the fricative constriction.

2.3.1. Image Analysis

The midsagittal slice was selected from each stack and cropped in Matlab to select the same region of interest across scans and subjects. Image contrasts were hyper-adjusted to enhance air-tissue boundaries. Palate, tongue and lip outlines were then located using an automatic edge-detection technique based on the Canny algorithm.

The midsagittal profiles extracted in this manner were superimposed to allow comparison of fricative articulations across sibilants and in each vocalic context.

Additionally, three-dimensional surface models of the vocal tract were constructed from the MRI slice stacks using Able 3D-Doctor software and used to compare the geometry of the tract around the oral constriction and the lips.

2.4. Acoustic Recordings

All fricatives produced by subjects during the MR Imaging were recorded using a Phone-Or ceramic noise-cancelling microphone system integrated into the scanner. In this manner the veracity of each fricative and vowel elicited during the scanning session could be verified, and the quality of the fricative production could be monitored.

In order to obtain high-fidelity recordings of the fricatives for acoustic analysis, each subject was recorded in an anechoic recording studio at Haskins Laboratories during a separate session. The same fricatives that were produced in the scanner were elicited and recorded during a 90 minute recording session. Subjects were seated on a chair in the centre of studio and asked to read timed IPA prompts presented using EPrime software on a laptop positioned at the far end of the room. A Brüel&Kjær 4190 Far Field Microphone was located 650mm in front of the lips of the subject, 25° off-centre. All data were recorded digitally (16 bits, 44100Hz) using a custom data acquisition device operating on National Instruments hardware.

2.4.1. Acoustic Analysis

A significant amount of low frequency (<30Hz) ambient noise was detected in the recording environment, so all recordings were high-pass filtered through a third-order elliptical IIR filter with a 50Hz passband.

A 1.4sec segment was selected from each recording for spectral analysis, taken 2.5sec after the beginning of frication in each token. The power spectral density for each of the fricatives recorded was estimated from this steady-state segment with the Welch method, using 1024 point FFTs calculated over 50% overlapping 1024 point Hamming windows, and plotted on a linear scale from DC to the Nyquist frequency 22050Hz. PSDs of each fricative/vowel combination were calculated for all five subjects and superimposed for comparison.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after abstract only review

3. Results

3.1. Contrastive Production of Sibilants

Numerous articulatory differences were evident across the five subjects in the production of each sibilant. However the MR images revealed less variation in the production of [s] than for [ʃ].

3.1.1. Articulation of the Alveolar Sibilant

Despite individual differences in production identified below, collectively the images suggest that articulation of the alveolar fricative [s] was characterised by:

- a narrow constriction (often too narrow to be resolved) formed with the tongue blade in the dental-alveolar region
- absence of a sublingual cavity, or a sublingual region too small to be identified

The place of articulation ranged from dental to near post-alveolar. Midsagittal images of the most dental (subject F2), and the most posterior (subject M2) productions of [s] are shown in Figs. 1 and 2 respectively.

Figure 1: Dental articulation of [isːi] (subject F2)

Figure 2: Posterior articulation of [usːu] (subject M2)

3.1.2. Articulation of the Post-alveolar Sibilant

Considerably more variation was evident in the articulation of [ʃ] than for [s] across this population of speakers. Four of the five subjects produced [ʃ] with a raised tongue blade forming a distributed constriction extending from the palatal to the post-alveolar region, while one subject (F3) produced [ʃ] with a slightly retroflexed tongue posture which produced a less distributed constriction. The two configurations are contrasted in Figures 3 and 4.

Figure 3: Distributed articulation of [uʃːu] (subject M2)

Figure 4: 'Retroflexed' articulation of [aʃːa] (subj. F3)

These figures also demonstrate two other differences in articulation evident in the production of [ʃ]:

- the size of the sublingual cavity, which ranged from non-evident (subject F1) to highly prominent (subject F3)
- the amount of lip-rounding used

Differences in lip rounding are best revealed through the use of superimposed midsagittal outline images. The two fricative articulations superimposed in Fig. 5 show that for subject M1, lip rounding and lip protrusion are clearly evident in the production of [ʃ], but are not employed in the articulation of [s]. Fig. 6 shows that subject F3, on the other

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

hand, does not use differential labial articulations to produce her sibilants.

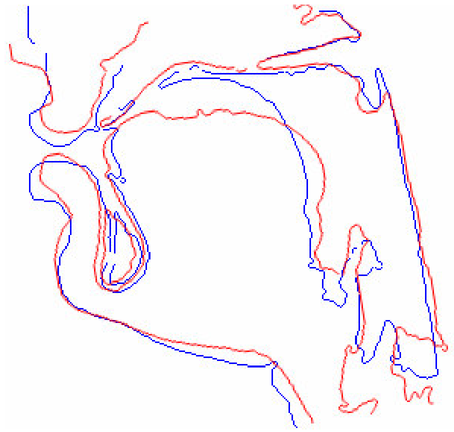

Figure 5: Comparison of lip-rounding [s-ʃ] (subj. M1): red=[us:u], blue=[uʃ:u]

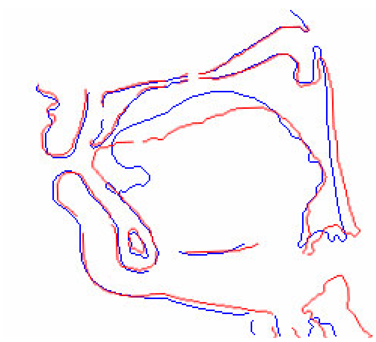

Figure 6: No rounded articulation for [ʃ] (subject F3): red=[us:u], blue=[uʃ:u]

3.2. Influence of Vocalic Context on Sibilants

Differences in articulation due to vocalic context were evident for all subjects, although vocalic effects were more noticeable for some subjects than for others.

Some of the differences in articulation which have already been identified are either partially attributable to vocalic context, or enhanced by vocalic influence. While all of subject F3's post-alveolar fricatives are less distributed than those of the other subjects, for example, the most extreme difference is observed in the token [aʃ:a], where the lowered jaw accentuates the retroflexed posture of the tongue (Fig. 4). Likewise, although every [s] produced by subject F2 is more anterior than those produced by subject M2, the tokens illustrated in Figs. 1 and 2 are the two most extremely differentiated fricatives because of the contrasting vocalic contexts (front vs. back).

In Fig. 7, midsagittal images of the articulation of the alveolar fricative [s] as uttered in three different vocalic contexts have been superimposed. For this subject (F2), the only observable difference is a slight retroflexion of the tongue in the context of the low vowel.

Figure 7: Effect of vowel on [s] (subject F2): red=[is:i], yellow=[as:a], brown=[us:u]

Fig. 8, in contrast, reveals that for subject M2, the three different vocalic contexts produce three quite different articulations of [ʃ]. Both jaw and lips are positioned differently for all three tokens. In the context of the low vowel, the tongue shape differs enough to affect the length of the constriction, the shape of the tongue dorsum, and the size of the sublingual cavity.

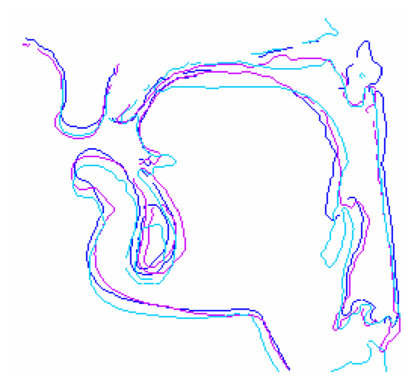

Figure 8: Effect of vowel on [ʃ] (subject M2): blue=[iʃ:i], aqua=[aʃ:a], purple=[uʃ:u]

Articulatory differences due to vocalic context may also be demonstrated by comparing 3D models of the labial region of the face. The surface models were constructed from the sagittally-oriented MRI image stacks. By aligning the models using the surface of the chin as a landmark, the relative protrusion of the lips can be compared for two different utterances. It can be seen in Fig. 9, for example, that subject M1 produces the alveolar sibilant [s] with significantly greater lip protrusion in the context of a rounded vowel ([us:u]) than for a spread vowel ([is:i]).

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

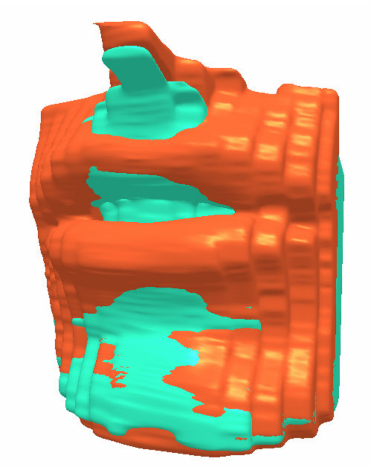

Figure 9: Comparison of labial articulations of [s] (subject M1): aqua=[is:i], orange=[us:u]

In contrast, in Fig. 10 we can see that for the same subject, the post-alveolar sibilant [ʃ] is produced with nearly identical lip rounding and protrusion in the same two vocalic contexts.

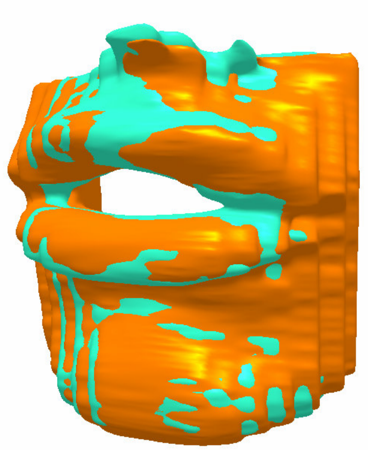

Figure 10: Comparison of labial articulations of [ʃ] (subject M1): aqua=[iʃ:i], orange=[uʃ:u]

3.3. Acoustic Correlates to Articulatory Variation

Much of the articulatory variation observed here resulted in acoustic variation which may be readily identified in the spectra of the fricatives. In Fig. 11, the spectra of the two alveolar sibilants which displayed the greatest articulatory difference are compared. The alveolar [s] produced by subject M2 is characterised by a strong peak around 3kHz, while the dental [s] produced by subject F2 shows a peak around 7.2kHz, with the bulk of the spectral energy distributed broadly above 5kHz.

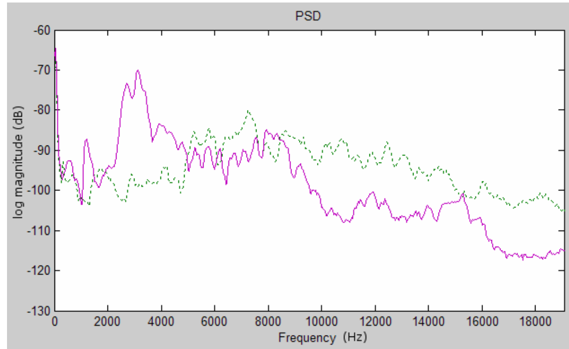

Figure 11: Comparison of [s] spectra: solid=[us:u](subj.M2), broken=[is:i](subj.F2)

This effect is not restricted to tokens taken from different subjects. With some subjects, vocalic context effects are sufficiently pronounced that the spectra of the two sibilants begin to converge. In Fig. 12, two spectra for subject M2 are compared: that of the alveolar fricative in the context of a back rounded vowel, and the post-alveolar fricative in the context of a high front vowel. The spectra are remarkably similar, both displaying a strong peak around 3kHz, indicating the magnitude of the effect of vowel context on sibilant quality for this speaker.

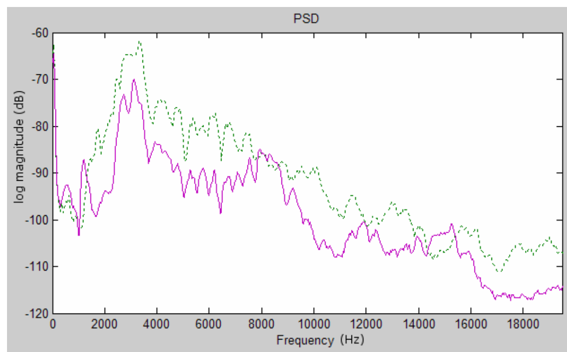

Figure 12: [s-ʃ] spectra – effect of vocalic context: solid=[us:u](subj.M2), broken=[iʃ:i](subj.M2)

The two subjects whose sublingual cavities were most prominent in the articulation of [ʃ] (M2, F3) both produced spectra with prominent narrow spectral maxima in the vicinity of 2kHz. Four of these six tokens were characterized as having a whistly component. Although most of the [ʃ] spectra also displayed strong components in the band below 4kHz, the spectra of the sibilants produced with a sublingual cavity tended to have a narrower (<400Hz) single spectral maximum in this band, best illustrated in the tokens [uʃ:u] (Fig. 13).

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after abstract only review

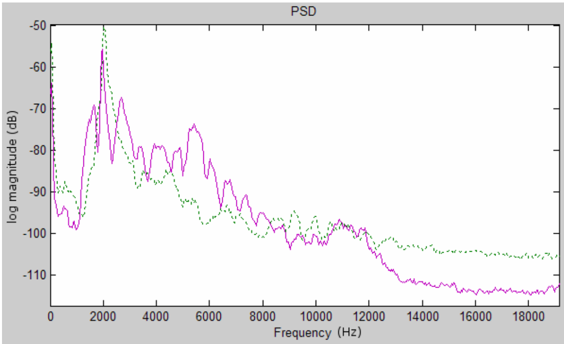

Figure 13: [ʃ] spectra – tokens produced with sublingual cavities: solid=[uʃːu](subj.M2), broken=[uʃːu](subj.F3)

4. Discussion

While this study has considered some of the most important articulatory features of sibilants, these are only a few of the many dimensions which must be considered in a thorough characterisation of sibilant articulation. Midsagittal analysis provides important insights into the basic configuration of the tongue and lips during fricative articulation; however it is known that other variables, such as the geometries of the constriction and the front cavity, have important acoustical effects (Shadle, 1991). Better characterisations of sibilants will result from 3D analysis of the geometries in these vicinities, and from the analysis of MR image orientations other than sagittal. This is an area of ongoing research.

5. Conclusions

The following factors were found to be subject to articulatory variation in the production of sibilants for this population of speakers:

- the location of the constriction of both [s] and [ʃ]
- the posture of the tongue with respect to the palate in the articulation of [ʃ]
- the extent of distribution of the constriction for [ʃ]
- the size of the sublingual cavity produced in the articulation of [ʃ]
- the amount of lip rounding employed in the articulation of [ʃ]

Individual differences between speakers, and vocalic context effects were both sources of variation for each of these factors. Much of this articulatory variation corresponded to significant acoustic variation between the fricatives produced. The acoustic properties of [s] in particular varied between speakers and as a result of vocalic influences.

The results suggest that neither the articulatory nor the acoustic properties of sibilants can be considered to be invariant, and that the production of fricatives seems to involve the complex interaction of both articulatory and acoustic factors.

6. Acknowledgements

This work is part of an ongoing project supported by NIH grant NIDCD R01 DC 006705.

7. References

Browman, C. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6: 201-251

Cohen, M. & Perkell, J (1986). Palatographic and acoustic measurements of the fricative consonant pair /s/ and /š/. *Proc. 12th Int. Congress on Acoustics*, Toronto, A3-5

Fowler, C.A., Rubin, P., Remez, R.E., & Turvey, M.T (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language Production* 373-420. New York: Academic Press

Narayanan, S., Alwan, A. & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *JASA* 98: 1325-1347

Perkell, J.S., Boyce, SE. & Stevens, K.N. (1979). Articulatory and acoustic correlates of the [s-š] distinction, *Speech Communication Papers*, 97th Meeting of the ASA, J.J. Wolf and D.H. Klatt (eds.), 109-113

Shadle, C.H. (1991) The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics* 19: 409-424

Shadle, C. & Scully, C. (1995). An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics* 23: 35-66

Shadle, C., Mair, S., Carter, J. & Millner, N. (1995). The effect of vowel context on acoustic characteristics of [ç, x]. *Proc. XIIIth International Congress of Phonetic Sciences*, Stockholm 1: 66-69

Shadle, C., Tiede, M., Masaki, S., Shimada, Y. & Fujimoto, I. (1996). Articulatory effects of vowel context on fricatives: an MRI study. *Proc. Institute of Acoustics* 18:9:1, 187-194

Stevens, K. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press

Accepted after abstract only review