INTER-SPEAKER VARIABILITY IN SIBILANT PRODUCTION AND SOUND CHANGE INVOLVING SIBILANTS*

Alice Faber

Haskins Laboratories, New Haven, CT 06511 USA

ABSTRACT

The role of contextual and inherent lip rounding in mediating synchronic and diachronic interchanges between [s] and [ʃ] is described in the context of a model of sound change involving inter-speaker variability in speech production and perception strategies.

1. INTRODUCTION

In a variety of the world's languages, interchanges of [s] and [ʃ] in labial environments, that is, near sounds like [r m u ʕ (ɨ)], are attested. Such changes are seen in some ancient Semitic languages [3], Tigriña (a modern northern Ethiopian Semitic language) [12], early Indo-European [7], and southern American English [9]. Based on published studies showing that [ʃ] is rounded in a variety of languages [1,2,5], that [s] is lower frequency adjacent to [u] than to other vowels [4], and that the frequency boundary between [s] and [ʃ] is higher before [u] than before [a] [6], I attributed the changes in both directions to the phonologically ambiguous status of a phonetically rounded sibilant in a rounded context: if the lowered frequency is attributed to the rounded context, the sibilant is interpreted as [s]; but, if, instead, the lowered frequency is interpreted as inherent to the sibilant, the sibilant is interpreted as [ʃ] [3]. Sound change, in this model, is a result of individual differences in speech production and perception: Speakers will differ in how much rounding, inherent or contextual, they produce in a given instance, and listeners will differ in their interpretation of rounding in a specific instance as inherent or contextual.[1]

This paper is a tentative report of a preliminary series of experiments aimed at testing the hypothesis regarding [s]-[ʃ] interchanges as well as the general model of sound change in which that hypothesis is embedded. The present hypothesis is that [s] and [ʃ] will be less distinct acoustically in some labial contexts than in non-labial contexts for at least some speakers of some languages. Furthermore, this decreased distinctiveness should result in part from increased rounding of [s] in these labial contexts.

2. METHODS

There were 7 subjects for these experiments, including one Polish-English bilingual. All speakers produced utterances of the form *VSV*, with the flanking vowel {a i u}, and S one of {s ʃ} (vowel tokens), and the bilingual speaker in her Polish mode produced utterances with [ʂ ɕ]; the utterances with [ʂ] will not be discussed here, and [ɕ] will be treated as equivalent to [ʃ]. Two speakers also produced *aC(a)Sa* and *aS(a)Ca*, utterances with S again one of {s ʃ}, and C one of {k l m r} (consonant tokens). All tokens had penultimate stress. Subjects, their language backgrounds, and their data are summarized in Table 1.

Lip position was monitored with a modified Selspot opto-electronic tracking system. For some subjects, linguo-palatal constriction location was monitored with a RION artificial electropalate. Movement signals were digitized at 200

Table 1: Summary of subjects and data. Under Data Set, *vowel* means *VSV* utterances, and *cons* *aC(a)Sa* and *aS(a)Ca* utterances.

Subject	Language	Data Set	# of tokens
EF	Italian	vowel	50
JM	German	vowel	50
DR	Catalan/Spanish	vowel	50
ED	Polish	vowel	10
	English	vowel	10
EVB	English	vowel	40
KSH	English	vowel + cons	10
FBB	English	vowel + cons	10

samples/sec., and the EPG signal at 64.1 frames/sec. The speech signal, recorded on a Telex unidirectional head-set microphone, was digitized at 20,000 samples/sec. (12 bits), without preemphasis. Automatic peak picking algorithms were used to identify upper lip protrusion (ULP) maxima and minima; for segments without clear extremes, an arbitrary point was measured. For present purposes, the measure of sibilant frequency used was the Centroid. Centroids were computed for all sibilants over the range c. 1,000–10,000 Hz.[2] In vowel tokens, one centroid was computed, at the approximate midpoint of S. For consonant tokens, three centroids were calculated, one in the middle, and one at each edge.

For each speaker, the extent of variation in sibilant frequency and in ULP was assessed via Analyses of Variance. For vowel conditions, the analysis was Vowel X Sibilant, and for consonant conditions, Consonant X Sibilant X Order X Adjacency, with Measurement Point as a repeated factor for the acoustic ANOVA. Order refers to whether C preceded or followed S, and Adjacency to whether S and C were abutting or separated by a stressed [a].

3. RESULTS AND DISCUSSION

3.1 Acoustic Factors

All vowel condition speakers had significant main effects of Sibilant and Vowel, as well as significant interactions. Centroids were lower for [ʃ] than for [s], and were lower when the flanking vowel was [u] than in the other contexts. Fig. 1 shows the extent to which [s] and [ʃ] were distinct in the three vowel contexts.[3] The higher the percentage, the larger the frequency difference between [s] and [ʃ]. There are clear differences among subjects in the extent to which they distinguish the two sibilants, as well as in the nature of the vocalic effects on the distinction. The primary commonality is that the two sibilants are less distinct in the [u] context than in the other vocalic contexts. Examination of the raw centroid values reveals, further, that the decrease in sibilant distinctiveness in the [u] context results primarily from a decrease in frequency for [s]. Decreases in distinctiveness in the [i] context result, in contrast, from an increase in frequency for [ʃ].

Analyses for the consonant subjects are, not surprisingly, more complicated.

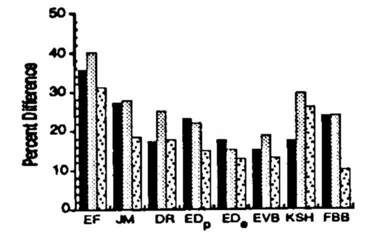

Figure 1: The percent difference between [s] and [ʃ] in mean centroid frequency for eight subjects. Dark bars are values for the [i] context, light bars for the [a] context, and speckled bars for the [u] context.

Both had significant main effects for Sibilant, Consonant, Order, Adjacency, and Measurement Point, as well as many interactions. The latter result is not surprising, since differential effects of context consonants on sibilants should be strongest at the sibilant edge adjacent to the context. In any case, [s] was higher frequency than [ʃ]. For KSH, sibilants were lower cooccurring with [r m] than with [k l], while for FBB, sibilants in tokens with [r l] were lower than those with [k m]. For KSH, Consonant effects on sibilants were stronger for [s] than for [ʃ], while for FBB the reverse was true. For both subjects, the sibilant was lower frequency at the third measurement location than at the earlier two.

The frequency differences between [s] and [ʃ] in the different consonantal contexts are illustrated in Figs. 2 and 3 In brief, these figures show that [s] and [ʃ] are slightly less distinct in the bisyllabic forms (in which S and C abut) than they are in the trisyllabic forms. For both subjects, one of the largest distinctions between the sibilants occurs, surprisingly, *after* [r]. For FBB, but not for KSH, there is a marked decrease in sibilant

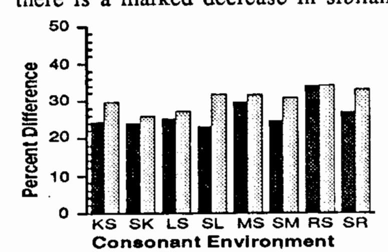

Figure 2: The percent difference in mean centroid frequency between [s] and [ʃ] for subject KSH. The *x*-axis legend shows the relative order of context consonant and sibilant. Dark bars represent sibilants adjacent to the consonant, and light bars sibilants separated from it by a stressed vowel.

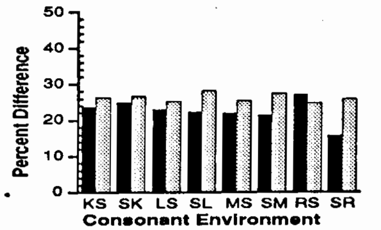

Figure 3: The percent difference in mean centroid frequency between [s] and [ʃ] for subject FBB. See Figure 2 caption for further details.

distinctiveness before [r]. As Fig. 1 shows, she also has a clear decrease in sibilant distinctiveness in the [u] context.

3.2 Articulatory Factors

While it is customary to attribute differences in sibilant acoustics to differences in linguo-palatal constriction (a more anterior constriction producing a higher frequency sibilant [10, 11]), we will first consider the effects of differential ULP on the acoustic differences observed so far. For all subjects, there were significant main effects of Sibilant and Vowel. There was consistently more ULP for [ʃ] than for [s], and more for sibilants next to [u] than for those next to [i] or [a]. As is clear from Fig. 4, which shows the percentage difference between normalized[4] ULP for [s] and [ʃ], subjects differ markedly in the extent to which they have a ULP difference between [s] and [ʃ] and in the extent to which the difference is sensitive to vocalic context. Subjects who, like JM, have a ULP contrast between [s] and [ʃ] in a rounded context, have more protrusion for [ʃ] in such a context than in an unrounded context. Other subjects, like EF, neutralize the protrusion contrast between [s] and [ʃ] in a rounded context.

The two consonant subjects had somewhat different patterns of ULP. Subject KSH, as shown in Fig. 5, has less con-

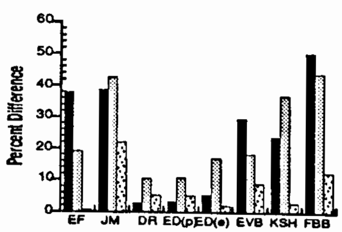

Figure 4: The percent difference in mean upper lip protrusion between [s] and [ʃ] for 8 subjects. See Figure 1 caption for further details.

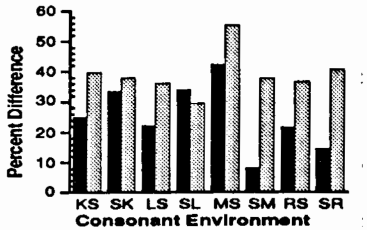

Figure 5: The percent difference in peak lip protrusion between [s] and [ʃ] for subject KSH. Conventions are as in Figure 2.

trast in ULP in the bisyllabic tokens than in the trisyllabic ones. She further has less contrast in syllable initial sibilants than in syllable final ones, especially before [m] and [r], the consonants most likely to induce ULP in a preceding [s]. In contrast, FBB has more ULP contrast in trisyllables than in bisyllables, as shown in Fig. 6. This pattern makes it all the more striking that she, too, has much less contrast between the sibilants immediately preceding [m] or [r] than in the other contexts.

In some instances, there is a clear relationship between the acoustic and labial patterns of variation. These are, for the most part, instances in which the context segment is labial. For 6 of the 8 vowel condition subjects, the decrease in magnitude of the acoustic difference between [s] and [ʃ] is commensurate with the decreased distinction in ULP. However, two subjects, EF and KSH, have nearly as much acoustic difference between the two sibilants in the [u] context as in the other contexts, but virtually no difference in ULP. Since these are two subjects for whom EPG data are available, the linguo-palatal basis for the distinction can be identified. For KSH, both [s] and [ʃ] are retracted in the [u] context, and both are, therefore, lower frequency; the distinction is thus preserved.

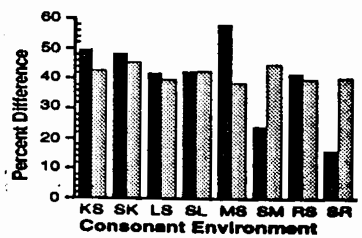

Figure 6: The percent difference in peak lip protrusion between [s] and [ʃ] for subject FBB. Conventions are as in Figure 2.

EF, in contrast, has a more *anterior* [s] in the [u] context than elsewhere, thus counteracting the frequency-lowering effects of the increased ULP.

The two consonant subjects have congruent acoustic and labial variation patterns in utterances containing [m r], but less congruent ones in utterances with [k l]. This is sensible, given that the former contribute active ULP gestures while the latter do not. Unfortunately, no EPG data are available for FBB. However, interpretation of KSH's EPG data is straightforward here too. First, [s] is retracted following [r], while [ʃ] is not affected by which side of [r] it is on; this decreased distinctiveness reinforces the effects of the decrease in ULP. Fig. 5 shows an additional pattern of differential effects of preceding and following [k] and [l] on ULP distinctiveness, a pattern not reflected in acoustic differences: [s] and [ʃ] are less distinct in ULP following [k] or [l] than preceding them. However, the EPG data show the reverse: [s] and [ʃ] are more distinct in linguo-palatal constriction location following [k] or [l] than preceding them. The effects of this linguo-palatal difference apparently cancel out those of the ULP difference.

4. CONCLUSION

The experimental results described above show that the decreased acoustic distinctiveness of sibilants in rounded contexts is the result of a complex interplay of labial and linguo-palatal factors. Speakers vary in the relative contributions of the two sorts of factors to their acoustic patterning. In order to preserve the acoustic invariance of [s] across a range of contexts, speakers must vary their linguo-palatal targets for [s], in an attempt to compensate for context-based variation in lip position. In contrast, preservation (to the extent possible) of articulatory invariance, at least as regards linguo-palatal constriction, leads to an increase in acoustic variability. Concentration on articulatory invariance can lead to instances of [s] that might be perceived as [ʃ]. But, concentration on acoustic invariance can lead to a proliferation of articulatory targets, the number of which may subsequently be reduced in novel ways, especially the reinterpretation of some instances of [ʃ] as [s]. Either way, sound change occurs.

5. REFERENCES

[1] ABRY, C. & J.-L. BOE, 1986, "Laws for Lips", *Speech Communication*, *5*, 97–104.
[2] BROWN, G., 1981, "Consonant Rounding in British English", *Towards a History of Phonetics*, R. E. Asher & E. J. A. Henderson, eds. Edinburgh: University of Edinburgh Press, pp. 67–76.
[3] FABER, A., 1986, "On the Actuation of Sound Change", *Diachronica*, *3*, 163–184.
[4] HEINZ, J. M. & K. N. STEVENS, 1961, "On the Properties of Voiceless Fricative Consonants", *JASA*, *33*, 589–596.
[5] MALMBERG, B., 1963, *Phonetics*, NY: Dover Press.
[6] MANN, V. A. & B. H. REPP, 1980, "Influence of Vocalic Context on Perception of the [s]-[ʃ] Distinction", *P&P*, *28*, 213–228.
[7] MARTINET, A., 1955, *Économie des changements phonétiques*, Berne: A. Francke.
[8] OHALA, J. J., 1989, "Sound Change is Drawn from a Pool of Synchronic Variation", *Language Change*, L. E. Breivik & E. H. Jahr, eds. Berlin: Mouton de Gruyter, 173–198.
[9] PENNINGTON, M. C., 1982. *The Story of S*, Philadelphia: University of Pennsylvania Ph.D. Dissertation.
[10] STEVENS, K. N., 1972, "The Quantal Nature of Speech", *Human Communication*, E. David & P. Denes, eds. NY: McGraw Hill, 51–66.
[11] STEVENS, K. N., 1989, "On the Quantal Nature of Speech", *J. Phon.*, *17*, 3–45.
[12] VOIGHT, R. M., 1988, "Labialization and the So-Called Sibilant Anomaly in Tigrinya", *BSOAS*, *60*, 525–536.

6. NOTES

* Research support from NIH grant DC-00016 is gratefully acknowledged.

[1]This view of sound change is like that of Ohala [e.g., 8]. The primary difference is that Ohala sees listeners' varying interpretations as *errors*, while I attribute them to inherent indeterminacies in the process of speech perception.

[2]This range was selected to minimize the influence of partial voicing of some tokens. In addition, there was a small amount of relatively low frequency interference as a result of the recording set-up, and it was necessary to reduce the impact of this interference on the centroids.

[3]Numbers for this and following figures were derived by the formula $n = 100 * \frac{x_1 - x_2}{x_1}$ with x_1 the larger of the two inputs.

[4]Normalization took place in two steps. During the analysis for each subject, raw ULP values were normalized with reference to a fixed rest position. Later, the lowest mean value for each subject was set to 5 mm, so that the variability apparent in Figure 4 reflects differences in subjects' articulatory patterns rather than scaling.