The interaction between prosodic boundaries and accent in the production of sibilants

Khalil Iskarous[1], *Marianne Pouplier*[1,2], *Stefania Marin*[2], *Jonathan Harrington*[2]

[1]Haskins Laboratories, New Haven, Connecticut, United States
[2]Institut für Phonetik und Sprachverarbeitung, LMU Muenchen, Germany
iskarous@haskins.yale.edu

Abstract

It is well-established that prosodic structure has an influence on speech production. However, a great deal of the work showing the influence of prosody on articulation and acoustics has focused on segments known to exhibit considerable variability in their production. Sibilants are highly constrained speech segments, due to the precise aerodynamic tasks they require. The goal of this work is to examine if boundaries of prosodic domains and accents are able to affect the production of sibilants. This study presents data from 5 subjects using Electromagnetic Articulography (EMMA), using a repetitive rhythmic speech task, where the repeated unit is composed of trochaic or iambic pairs. The results show that boundaries, accent, and rhythm do show effects on the magnitude of motion of articulators during sibilants.

Index Terms: prosodic boundaries, accent, speech production, sibilants

1. Introduction

This work concerns the relation between the two aspects of prosody, grouping and prominence [1, 2], and how these two aspects interact in the *rhythmic* organization of speech production. The segments we investigate are perhaps the ones least likely to show the effects of prosodic variation: sibilants. For the production of these segments, the articulators are highly constrained due to the difficulty of the achievement of their aerodynamic and motor tasks [3, 4]. Therefore, due to the high level of constraint upon the articulators, it may be expected that these segments would show little to no prosodic variability in their production. However, if prosody is not a low-level effect, but is rather the method for hierarchically grouping and organizing the contrastive units in speech production, then it should be possible to measure the effects of varying the two aspects of prosody, grouping and prominence, even on sibilant production.

Several works have shown effects of prosodic boundaries (an aspect of grouping) and prominence on stops [5, 6], but two studies that investigated the effects of prosody on sibilants showed marginal [7] or no significant effects [8] on spectral moments of the sibilant noise, which would be predicted to covary with gestural magnitude. These works did find some effects of prosody on the duration and overall energy in sibilants. Therefore there is support in the literature for the hypothesis that prosodic structure may affect properties such as duration, but not gestural magnitude, in highly constrained segments like sibilants. Using a rhythmic task, that allows for the investigation of grouping, prominence, and their manifestation in rhythm, we investigate the effect of these methods of hierarchical organization on the motion of the tongue and jaw in sibilants using Electomagnetic Articulography (EMMA). Specifically, we investigate how prosodic boundaries and accents, and their interaction, affects various properties of /s/ and /sh/ in American English.

2. Rhythm as Boundary x Accent

To motivate the rhythmic speech task used in this work, we use an explicit relation between rhythm, prosodic boundaries, and accent. This relation is implicit in [1]'s definition of *prosody* as "the grouping and relative prominence of the elements making up the speech signal. One reflex of prosody is the perceived rhythm of the speech". To investigate the effect of the grouping of contrastive units in a phrasal domain, specifically how initial vs. non-initial units in a phrase are produced, a factor Boundary can be introduced, with levels Initial vs. Medial, indicating that the segment is initial or non-initial in a phrase. And to investigate the effect of accent, as a form of prominence, on production, a factor Accent can be introduced, with levels Accented vs. Unaccented. It is possible to study the main effects of each of these factors, but it is also possible to study their interactions. If there are significant interactions of these two factors on production, it means that the behavior of a segment in two of the four possible combinations (Initial-Accented, Initial-Unaccented, Medial-Accented, Medial-Unaccented) behave in a way that is not predictable only from whether the segment is initial or not and Accented or Unaccented. For instance, it could be the case that Initial-Accented and Medial-Unaccented segments pattern together, and differently from Initial-Unaccented and Medial-Accented segments. But Initial-Accented and Medial-Unaccented are exactly what is present in a trochee, which sequences a strong beginning and a weak end. And Initial-Unaccented and Medial-Accented are exactly what is sequenced in an iamb. Therefore the presence of an interaction can be investigated through a rhythmic task that compares trochaic and iambic productions. This contrasts with another method for investigating interaction of grouping and prominence [6], where several utterances are used, and the same segment is produced in the four types of possible positions scattered through the sentences. In the rhythmic and repetitive speech task used here, the interaction between Boundary and Accent is made part of the task itself.

Several hypotheses are addressed in this work. One hypothesis is that prosody affects only segments that are not as highly constrained as the sibilants. A competing hypothesis is that Boundary would have an effect on articulator position, but that accent wouldn't. This hypothesis is based on previous findings of initial strengthening on closure magnitude for stops, but no such strengthening with accentuation [6]. Another hypothesis we test is that rhythm, as the interaction, of Boundary and Accent, also has an effect. If this turns out to be true, then this work would contribute to the investigation of speech rhythm. Besides the positions of articulators, results for duration and acoustic energy are also presented, in order to investigate whether the factors have differential effects on the various properties of the sibilants.

3. Methods

Articulatory data were collected using the 2D EMMA system [9] at Haskins Laboratories. Sensors were attached to nose ridge, maxilla, upper and lower lips, tongue tip (TT, about 1cm behind the actual tip), tongue blade (TB), tongue dorsum, and, to record jaw movement, to the lower incisors. The kinematic data were collected at 500 Hz and low-pass filtered with 15Hz, the acoustic data were collected synchronously with the kinematic data with a sampling rate of 20kHz. Standard calibration and post-processing techniques were performed for each experiment; see [10] for full technical details experimental setup and procedures. Subjects were instructed to repeat the phrases "sop sop" and "shop shop" in synchrony to a metronome beat for 10 seconds per trial. Two conditions were employed: iamb ("sop SOP") and trochee ("SOP sop"). During each trial, subjects saw the two-word phrase printed on the screen and accent placement was indicated to the subjects by capitalization. The trials are part of a corpus on metronome-paced speech collected for a different experiment. Subjects were instructed to align the stressed syllable to the metronome beat. Metronome rate was set to 120bpm, allowing for a speaker-specific adjustment within a +/- 4 beats per minute range of the target rate. The speaker-specific rate was determined during the practice trials. Auditory evaluation by two of the experimenters (Pouplier, Marin) confirmed that subjects produced the alternating stress patterns in the two conditions as instructed. Data from five native speakers of American English are included in the present analysis.

For each utterance, the sibilant portion in the acoustic signal was labelled semi-automatically. The positions of the articulators were extracted automatically at the mid-point of the fricative. The present analysis is focused on the vertical and horizontal components of the tongue tip, blade, and jaw, along with the root mean square energy (RMS) of a 20 ms window at the middle of the frication, and fricative duration. The jaw's effect on the tongue tip and blade were subtracted from the tip and blade measurements to reveal active motion of the tip and blade, rather than motion due to the jaw.

4. Results

Before presenting the results on the effect of prosody on sibilant production, we provide some baseline data on how /s/ and /sh/ are contrasted on the particular measures of production used in this work. The effects of prosody on the two sibilants will then be interpreted in light of the articulatory differences between them. A general linear mixed model was used to compare the effect of the contrast between /s/ and /sh/ on the various measures. The independent variable was PlaceArtic with two levels (Alveolar and PostAlveolar), and Subject was the random effect. Markov Chain Monte Carlo methods were used to compute the 95% confidence intervals for the effects and their significance at the .05 level [11]. Table 1 shows the effects of PlaceArtic. An empty entry indicates a non-significant difference or a significant difference below 1 mm, the resolution of EMMA.

Table 1: Effects of Place of Articulation on speech production measures. An empty entry indicates no significant result. The effects are in bold and surrounded by the lower and upper bounds of the 95% confidence intervals J = Jaw, TT = Tongue Tip, TB = Tongue Blade, X = horizontal component, Y = vertical component.

	Place of Articulation Effect
Duration	
RMS	
JY	
TTY	/sh/ Superior to /s/ [4.07,**4.42**mm,4.77]
TBY	/sh/ Superior to /s/ [9.47,**9.78**mm,10.1]
JX	
TTX	/sh/ Posterior to /s/ [4.08,**4.53**mm,5.01]
TBX	/sh/ Posterior to /s/ [2.19,**2.93**mm,3.69]

Table 2: Effects of Boundary, Accent, and Rhythm (interaction of Boundary (Initial vs. Medial) and Accent (Accented vs. UnAccented)) on speech production measures for /s/ and /sh/. An empty entry indicates no significant result. The direction of the effect is on the first line of the cell, the magnitude of the effect is on the second line, and the lower and upper bounds of the confidence intervals are on the third line. Duration is in ms units, RMS is in arbitrary units, and articulator motions are in mm units.

	Boundary		Accent		Rhythm	
	/s/	/sh/	/s/	/sh/	/s/	/sh/
Dur	I>M 13 [8,18]	I>M 22 [17,26]		A>UA 18 [13,23]		Tr > Im 19 [13,25]
RMS		I>M 38 [25,52]	A > UA 46 [30,62]	A > UA 54 [40,68]	Tr > Im 67 [45,89]	Tr > Im 58 [38,77]
JY			UA > A 1.24 [0.99,1.48]		Im > Tr 2.06 [1.72,2.41]	
TTY			A > UA 1.18 [0.93,1.42]	A >UA 1.25 [.84,1.62]	Tr > Im 2.22 [1.87,2.58]	Tr>Im 2.09 [1.55,2.63]
TBY						Tr>Im 1.04 [.38,1.66]
JX TTX TBX						

Place of articulation has no significant effect on the duration or level of acoustic energy (RMS) of the sibilants. Also, the jaw configuration, in both the horizontal and vertical dimensions, were not affected by Place of Articulation. However, the placement of the tongue tip and blade is significantly more superior and posterior for /sh/ than /s/.

In order to investigate the effect of the edge of the prosodic domain and accent, and their interaction, on the production measures, a general linear mixed model was fitted to the data using two independent variables Boundary (Levels: Initial (I) and Medial (M)) and Accent (Accented (A) and UnAccented (UA)). The interaction between the two independent variables, which as we argued earlier is Rhythm, was also calculated. The random effect was Subject. 95% confidence intervals and significance of the effects were computed as in the earlier analysis.

Table 2 presents the results of the statistical analysis. Boundary has a main effect on the duration of /s/ and /sh/, with the sibilant at the beginning of the phrase being longer than the phrase-medial sibilant. Accent has no main effect on /s/, but affects /sh/, with the accented /sh/ being longer than the weak /sh/. There is no significant interaction between the two independent variables for /s/, but for /sh/, the trochaic rhythm results in a significantly longer sibilant. Boundary has a significant effect on the RMS energy for /sh/, but not /s/. Again, the direction of the effect of Boundary is that the phrase-initial /sh/ has higher energy than the phrase medial /sh/. Accent has a significant effect on both /s/ and /sh/ RMS, with the sibilant in accented position having higher energy. And for both /s/ and /sh/, the trochaic rhythm has the effect of significantly increasing the energy of the sibilant.

It can be seen in Table 2 that the horizontal components of the articulators examined are not significantly affected by edges, accent, or rhythm. Furthermore, Boundary has no significant main effect on the vertical component of articulator position at the midpoint of the sibilant. An accented /s/ is lower, not higher, than a non-accented /s/. Furthermore, the iambic rhythm significantly raises the jaw for /s/. Neither accent nor rhythm affect /sh/ significantly. The active vertical component of the tongue tip (tongue tip y - jaw y), shows the opposite effect as the jaw's vertical motion, with Accented /s/ being higher than Unaccented /s/, and the trochaic /s/ raising the tongue tip. The same effect on the vertical component of the active tongue tip is also present for /sh/. The vertical component of the blade is affected by prosody in only one case, as an effect of the trochaic rhythm which raises the tongue body.

5. Discussion

Several generalizations emerge from the results, which allow us to address the predictions of the hypotheses of presented in Section 2. First, there is little support for the hypothesis that there is no effect of prosodic structure on sibilant production. Indeed, it can be seen from Table 2 that several aspects of sibilant production are indeed significantly affected. It is the pattern of the effects that is interesting.

Table 1 establishes, for the speakers that participated, a baseline of how /s/ and /sh/ are contrasted articulatorily. Duration, RMS, and Jaw position are not significantly different among the two sibilants. The differentiation is in the positioning of the tip and blade. The results on jaw and tongue support earlier findings [12]. One hypothesis based on these results is that prosodic variation would affect the aspects of production that do not distinguish between /s/ and /sh/, but would not be present in aspects that do distinguish between the two sibilants. This hypothesis is based on the idea that contrast acts as a Accented constraint on variability. The results do not support the strongest form of this hypothesis for all the effects, since there is variability in tip and blade position and there is no prosodic influence on JX, even though it does not significantly differentiate /s/ from /sh/. There is however support for a weaker version of the hypothesis, since the magnitude of the effect of prosody on tip and blade position is in the range of 10%-50% the magnitude of the difference between /s/ and /sh/ for these articulators. Therefore contrast does seem to play an important role in the magnitude of the prosodic effect. The importance of contrast is most clearly seen in the absence of an effect for the horizontal components of the articulator positions. The contrast between /s/ and /sh/ is in terms of the horizontal positioning of the constriction, rather than the vertical positioning of the tongue (e.g. in traditional feature specifications, /s/ is +anterior, while /sh/ is -anterior). It is articulatorily true that the tip and blade are more superior and posterior for /sh/ than /s/, as can be seen in Table 1, but the degree to which /sh/ is more superior is affected by prosody, but the contrastive horizontal positioning is highly immune to prosodic effects. This does support the hypothesis that contrast plays a major role in constraining variability. This is seen once the role of different articulators in the achievement of contrasts is clarified.

The direction of the effect of the factors and their interactions on Durations and RMS are consistent, when present: 1) Phrase-initial sibilants are longer and louder than medial sibilants; 2) Sibilants with the strong accent are longer and louder than weak sibilants; 3) The trochaic rhythm leads to longer and louder sibilants. The first two results are consistent with findings on other consonants The result on the strength of the trochaic rhythm could be explained by the predominance of trochaic patterns in English [13].

A surprising result of this study is that Accent, but not Boundary, had an effect on JY and TTY. This is surprising in light of earlier studies showing the opposite, that Boundary would have more of an effect than Accent on consonants [6]. Furthermore, Boundary has no effect on the articulators, even though Duration is lengthened in initial position (Table 2). Therefore the undershoot hypothesis relating longer segment time to higher chance of target achievement and shorter segment time with undershoot of target is not supported for Boundary, but is supported for Accent. One possible reason for this result is that since this data involved a fast speech rate, enforced through a metronome, there were not appreciable pauses between phrases, therefore the boundary strength could be low, thereby not having the expected influence.

The direction of effect for tongue tip and blade, when present, are the same for the main effects and interaction as for Duration and RMS. But JY shows the opposite effect, with Accented accent showing a lower jaw. This, however, supports the work of [14], which demonstrated that the jaw is *lower* under contrastive emphasis. It could be that the presence of an effect for prosody on articulator magnitude in this study, in contrast to the earlier studies of [8, 7], which used spectral center of gravity, could be due to the opposite effects of accent on JY and TTY. If JY is lowered, but active TTY is raised, then the constriction degree is maintained relatively constant across the prosodic contexts reducing an effect on the spectral center of gravity.

Moreover, the direction of effect for rhythm is the opposite for all the other measures, showing iambic sibilants having a higher jaw than trochaic ones. We believe that this effect is related to the lowering of the jaw for accented sibilants.

6. Conclusions

We have shown in this work that grouping and prominence do have significant effects on the production of sibilants, but that these effects are indeed limited by linguistic contrast. Moreover, Rhythm has an effect, as the interaction between Boundary and Accent.

7. Acknowledgements

This work was supported by NIH-NIDCD grant DC-02717, NIH-NIDCD grant DC-006705 , and DFG PO-1269/1-1.

8. References

[1] Pierrehumbert, J.,"Prosody and Intonation", in R. A. Wilson and F. Kell [Eds], The MIT Encyclopedia of the Cognitive Sciences, 479-482, Cambridge, MA: MIT Press, 1999.

[2] Keating, P.,"Phonetic Encoding of Prosodic Structure", in J. Harrington and M. Tabain [Eds], Speech production: Models, phonetic processes, and techniques, 167-186, Macquarie Monographs

in Cognitive Science, Psychology Press, New York and Hove, 2006.

[3] Hardcastle, W., *Physiology of Speech Production: An Introduction for Speech Scientists*, Academic Press,1976.

[4] Recasens, D. and Espinosa, A, "An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan", JASA, 125:2288–2298, 2009.

[5] Cho, T. and Keating, P., "Articulatory strengthening at the onset of prosodic domains in Korean", Journal of Phonetics, 28:155-190, 2001.

[6] Cho, T. and Keating, P., "Effects of initial position versus prominence in English", Journal of Phonetics, 37: 466-485, 2009.

[7] Silbert, N. and De Jong, K., "Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production", JASA, 123: 2769–2779, 2008.

[8] Cho, M. and McQueen, J., "Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress", Journal of Phonetics, 33:121–157, 2005.

[9] Perkell, J.S., Cohen, M.H., Svirsky, M.A., Matthies, M.L., Garabieta, I. and Jackson, M.T.T., "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements", JASA, 92: 3078–3096, 1992.

[10] Pouplier, M., Units of phonological encoding: Empirical evidence, Unpublished PhD dissertation, PhD dissertation, Yale University, 2003.

[11] Baayen, H., Analyzing linguistic data: A practical introduction to statistics, Cambridge University Press, 2007.

[12] Mooshammer, C., Hoole, P., and Geumann, A., "Jaw and Order", Language and Speech,50: 145–176, 2007.

[13] Cutler, A., and Carter, D. M., "The predominance of strong initial syllables in the English vocabulary", Computer Speech and Language, 2: 133–142, 1987.

[14] Erickson, D., Fujimura, E., and Pardo, B., "Articulatory Correlates of Prosodic Control: Emotion and Emphasis", Phonetica, 41: 399–417, 1998.