Evidence for the Dynamic Nature of Fricative Production: American English /s/

Khalil Iskarous, Christine H. Shadle, Michael Proctor

Haskins Laboratories, 300 George St., New Haven, CT 06511, USA
E-mail: iskarous@haskins.yale.edu

Abstract

This paper investigates whether sibilant consonants are essentially static segments, in which articulatory and acoustic variables assume a certain value at the beginning of the segment and maintain those values throughout the segment, with little articulatory and acoustic variability. Using the X-ray Microbeam database, analyzing /sVd/ productions for 24 speakers, we show instead that, consistent with other recent research, sibilants are dynamic segments that exhibit marked articulatory and acoustic variability through their time course.

1 Introduction

Sibilant fricatives are usually regarded as being among the most difficult segments to produce, due to the preciseness with which the vocal tract must be configured [1,2]. Successful sibilant production necessitates the formation of a constriction small enough to produce a jet, but large enough not to produce a stop. An obstacle must also be placed at a precise distance from the constriction to allow for aerodynamic-to-acoustic energy conversion characteristic of sibilants [3]. As a result of the preciseness with which sibilants must be produced, they are also highly resistant to coarticulation [4] and exhibit smaller spatial variability in their production than other fricatives [5]. Due to this high resistance to coarticulation and the reduced spatial variability, it is generally assumed that sibilants maintain relatively invariant articulatory configurations and spectral characteristics throughout their duration, i.e. the assumption is that sibilants are essentially *static* [1]. Moreover, according to Quantal Theory, even if there

is a small amount of articulatory motion during a sibilant, it will be acoustically hidden, since articulatory variability within the quantal region in which the sibilant is produced would cause little to no acoustic variability [6]. Under the hypothesis that sibilants are relatively static, one would then predict that the articulators most involved in their production, the tongue tip and blade and the jaw, would assume positions that allow the aerodynamic and acoustic consequences of their articulatory configuration to be achieved, and would then stay in that position until the end of frication, thereby resulting in minimal spatial and spectral variability within a sibilant.

Several recent works on sibilant production and acoustics provide data that allow us to question the traditional view of sibilants as static segments. Tasko and Westbury calculated the speed of jaw and tongue tip and blade pellets for sibilants and compared them to the speed of these articulators in other speech classes [7]. They found that the tongue tip/blade pellet is slower at the peak of the motion and at the boundary of the motion stroke for sibilants than other segments. The mandible pellet was also significantly slower than other segments at the boundary, but had a significantly *faster* peak speed inside the sibilant motion stroke. The latter result indicates significant dynamics of the jaw within a sibilant. A recent comparison of jaw motion in various coronal consonants found greater jaw motion for sibilants than other coronal segments [8]; however, it was not clear whether that is due to the dynamic nature of sibilant production or to the longer duration of sibilants. On the acoustic side, several studies using spectral moments of sibilants have provided data showing significant change in the first spectral moment, or center-of-gravity, of fricative spectra in the time course of a sibilant [9,10,11]. Moreover, Tabain has argued that first moment variability in

sibilants is greater than would be expected based on Quantal Theory [5].

In this paper, we directly investigate the variability of jaw motion and first spectral moment during the time course of /s/ to determine if these segments are essentially static or whether they exhibit an inherently dynamic component.

2 Methods

To investigate whether articulatory and acoustic characteristics change significantly through the time course of the segment /s/ in American English, we used the Wisconsin Xray Microbeam database (XRMB), since it allows us to study tongue, jaw, lips, and acoustic output with high temporal and spatial resolution for a large number of speakers. The XRMB database consists of the two-dimensional tracking of lips, tongue and jaw motion (40 Hz, but smoothed and upsampled to 145 Hz) with simultaneous recording of the speech wave (21700 Hz). Four pellets were glued on the tongue at regular distances to get an estimate of its functional divisions (the tongue tip (T1), blade (T2), dorsum (T3) and root (T4)). Two pellets were placed on the upper and lower lips and two additional reference markers on the jaw (at the central incisor and on a molar tooth). Details on the methods and subjects can be found in [17].

For this paper, jaw and the acoustic waveforms from task 13 were analyzed for 24 subjects. Task 13 consists of words of the form *sVd*, where *V* is one of ten American English vowels. The segmentation of /s/ was based on the onset and offset of high frequency aperiodic energy. For the articulatory analysis, nine evenly spaced samples of the Jaw displacement were obtained, with the first sample corresponding to the /s/ onset and the last sample corresponding to the offset, and the other samples evenly spaced in between. The jaw displacement measures were converted to jaw pitch angle using Westbury's method [18], in which the angle between the occlusal plane and the line extending from the lower-teeth pellet to the midsagittal projection of the molar pellet is computed. Note that the jaw pitch angle increases as jaw displacement decreases, that is, as the jaw opens more. For the acoustic analysis, nine 25-ms windows were spaced evenly through the segmented /s/ and a multitaper spectrum was

calculated for each window [12]. The windows overlapped by an amount inversely proportional to the duration of the /s/-token involved.

3 Results

To determine whether the jaw exhibits significant motion during an /s/, we calculated the range of pitch angle of the jaw during each token of /s/. In order to compare across subjects, the range during /s/ for each subject was then normalized by that subject's maximum jaw pitch angle during /ae/. The vowel /ae/ was chosen for the normalization since it tends to have the largest jaw opening, thereby providing a measure of the total range of motion of the jaw for the subject.
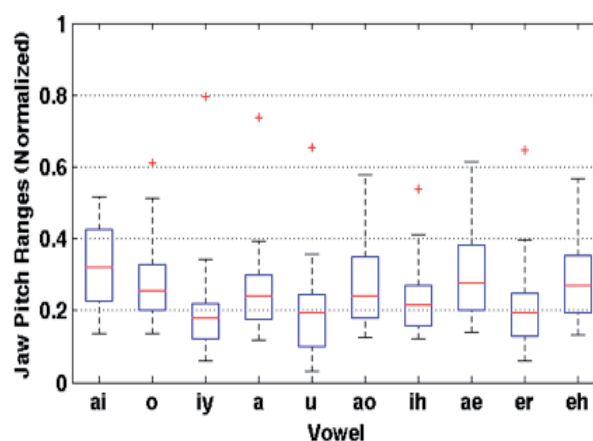

Figure 1: *Range of Jaw pitch angle during /s/ before 10 American English vowels, normalized within subject by the maximum jaw pitch angle during the vowel /ae/. The box-whisker symbols for each vowel context show data from 24 subjects. Each box represents the lower to upper quartile; red line in the box is the median. The whiskers show the minimum to maximum values. Crosses are outliers.*

The results of this analysis (see Figure 1) indicate that the jaw pitch angle ranges from about 0.2 to 0.4 of the total range of jaw pitch angle for each subject. This is a significant proportion of the total range of motion; therefore this result is incompatible with the notion that the jaw is relatively static during /s/. But the range of motion measure does not indicate the nature of the /s/ dynamic, only its magnitude.

To investigate the actual profile of motion, the trajectory of jaw pitch was visually examined for each token in the data and it was realized that there is always an upward, then a downward motion of the jaw during /s/ [13,14]. The time from frication onset to minimum jaw pitch during /s/ was then measured for each token. Figure 2 displays the histogram of these durations. It can be seen that only in a small minority of tokens does the jaw achieve its minimum opening at the beginning of the /s/. In the majority of tokens, the minimum jaw opening is not achieved until around the middle of the /s/. In a small minority of the tokens, minimum jaw opening is achieved late in the /s/.
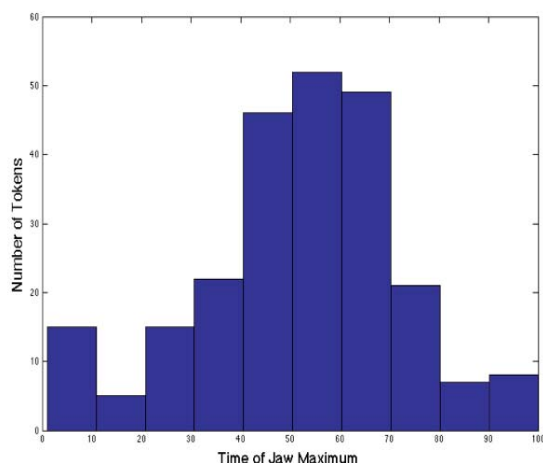

Figure 2: *Histogram showing distribution of the time during /s/ when jaw reaches its minimum jaw pitch angle (minimum jaw opening). Time = 0 is /s/ onset, time = 100% is /s/ offset. 240 tokens (24 subjects x 10 vowel contexts) are represented.*

As stated in the introduction, it could be that the motion of the jaw during /s/, however marked, is acoustically hidden. This would require articulatory motion to be within the quantal region for the segment /s/ in order to have little to no acoustic consequence, since ranges of the articulatory motion for segments are assumed to be chosen in such a way as to minimize acoustic variability. To examine this possibility, we computed the first spectral moment M1 through the time course of /s/ for each token. The spectral moment calculations used the multitaper spectra and then followed the method of Forrest et al. [15]. To be able to compare across subject, M1 calculated for each window in each token of each subject was normalized by subtracting the mean M1 calculated across all /s/ windows and tokens from that subject.
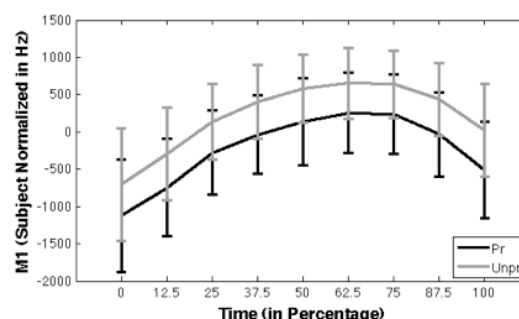

Figure 3: *Normalized M1 (first spectral moment) measurements for each time step of each token averaged across tokens, presented as means and standard deviations at each window location through /s/. Means are at the center of each error bar. Black line: /s/ tokens preceding rounded vowels. Grey line: /s/ before unrounded vowels.*

Results in Figure 3 are divided by whether /s/ precedes a rounded (black) or unrounded vowel (gray). Two effects can be seen: 1) /s/ preceding unrounded vowels has a higher M1 at each window location than /s/ preceding rounded vowels; 2) M1 shifts upwards and then downwards during /s/ regardless of whether the following vowel is rounded or unrounded. The first effect is possibly due to coproduction of the /s/ with the following vowel. If the labial gesture for a rounded vowel began early, it would lower M1, due either to a lengthened front cavity or decreased lip opening area, or both. The second effect could be due to a dynamic of the tongue tip, which is not studied here, or due to a change in glottal airflow, which may affect airflow at the sibilant constriction. However, it could also be due to the jaw dynamic presented in Figure 2. Raising the jaw during /s/ has the effect of raising the lower teeth, which allows them to act as more effective converter of aerodynamic to acoustic energy, raising the frequencies at which there is effective excitation, therefore increasing M1 as the jaw rises and decreasing it as the jaw falls. Whether M1 increases due to glottal airflow changes during /s/, tongue tip/blade motion, or jaw dynamic (or a combination of these), it can be seen that articulatory variability during /s/ has direct acoustic consequences. That is,

these acoustic consequences are not hidden, but are accessible to the listener.

4 Conclusion

Evidence was provided in this paper for the dynamic nature of sibilant production. This dynamic can be seen in systematic variability in articulatory and acoustic measures in the time course of the sibilant. This variability does not preclude preciseness in the production of sibilants, but does argue that preciseness does not necessitate static conditions. Sibilants are similar to other segments produced as the output of the overlap of gestural dynamical systems [16]. In future work, we aim to investigate the kinematics of other articulators like the tongue tip and dorsum, as well as spectral measures more complex than the first spectral moment, during the time course of sibilants.

Acknowledgement

References

[1] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Blackwell Publishers, Oxford, 1996.

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

[3] C.H. Shadle, *The Acoustics of Fricative Consonants*, PhD Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1985.

[4] D. Recasens, M.D. Pallares, and J. Fontdevila, A model of lingual coarticulation based on articulatory constraints, *Journal of the Acoustical Society of America*, (102) 544-561, 1997.

[5] M. Tabain, Variability in fricative production and spectra: implications for the Hyper- & Hypo- and Quantal theories of speech production, *Language and Speech*, (44) 58-93, 2001.

[6] K.N. Stevens, On the Quantal Nature of Speech, *Journal of Phonetics*, (17), 3-46.

[7] S.M. Tasko and J.R. Westbury, Defining and Measuring Speech Movement Events, *Journal of Speech, Language, and Hearing Research*, (45) 127-142, 2002.

[8] C. Mooshammer, P. Hoole, and A. Geumann, Jaw and Order, *Language and Speech*, (50) 145-176, 2007.

[9] A. Jongman, R. Wayland, and S. Wong, Acoustic Characteristics of American English Fricatives, *Journal of the Acoustical Society of America*, (108) 1252-1263, 2000.

[10] C.H. Shadle and S.J. Mair, Quantifying Spectral Characteristics of Fricatives, *ICSLP Proceedings*, 1521-1524, 1996.

[11] B. Munson, A method for studying variability in fricatives using dynamic measures of spectral mean, *Journal of the Acoustical Society of America*, (110) 1203-1206, 2001.

[12] O.S. Blacklock. *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods*. Ph.D. thesis, School of ECS, University of Southampton, 2004.

[13] R. McGowan, Anterior Tongue and Jaw Movement in sVd words, *Journal of the Acoustical Society of America*, (115) 2632, 2004.

[14] C.H. Shadle, K. Iskarous, and M. Proctor, Articulation of Fricatives: Evidence from X-ray microbeam data, *Journal of the Acoustical Society of America*, (119) 3301, 2004.

[15] K. Forrest, G. Weismer, P. Milenkovic, and R. N, Dougall, Statistical analysis of word initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* (84) 115–123, 1988.

[16] E. Saltzman and K. Munhall, A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, (1) 1615-1623, 1989.

[17] J.R. Westbury, *X-ray microbeam speech production database user's handbook*. Waisman Center, Madison: University of Wisconsin at Madison, 1994.

[18] J.R. Westbury, M.J. Lindstrom. Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research*, 45:651-662, 2002.