SIBILANT AND NON-SIBILANT FRICATIVES ARE PARSED ALIKE

Amanda Rysling and John Kingston

UC Santa Cruz and UMass Amherst
rysling@ucsc.edu, jkingston@linguist.umass.edu

ABSTRACT

Sibiliants /s/ and /ʃ/ concentrate energy at frequencies well above the rest of the speech signal. [1] contend that as a result sibilants stream auditorily in listeners' initial percepts. We present evidence from gating and categorization studies that non-sibiliant voiceless fricatives /f/ and /θ/ are segmented the same way that sibilant ones are in English. Because /f/ and /θ/ are broader in their energy distribution than sibilants, they overlap with context vowels and are less likely to stream auditorily than sibilants. It is the sharp acoustic discontinuity between voiceless noise and voiced periodicity, rather than spectral streaming, that causes listeners to parse fricative and vowel intervals as covarying, but clearly separate, intervals.

Keywords: Streaming, differentiation, prediction

1. INTRODUCTION

Many studies have shown that listeners respond "sh" more often to an /s-ʃ/ continuum before /i/ than /a/, /u/, or /y/ [2, 3, 4, 5, 6, 7, 9, 10]. This has been interpreted as compensation for coarticulation – listeners respond more often with the more posterior fricative /ʃ/ before the more anterior vowel – or as spectral contrast – they respond more often with spectrally lower /ʃ/ before the spectrally higher vowel. Listeners perceptually differentiate the fricative from the vowel, on either articulatory or auditory grounds. Listeners can also predict the quality of the following vowel from acoustic evidence in the fricative alone [5]. For /s/ and /ʃ/, both differentiation and prediction could be produced by auditory streaming [1]. Most, but not all, sibilants' energy is above that of the spectrally highest vowel, /i/. Streaming could separate the higher frequency energy of both /s/ and /ʃ/ from the lower portion of their spectra, where acoustic effects of coarticulation with the following vowel's F2 can be heard and used to predict that vowel's quality. If streaming is responsible for differentiation and prediction, then neither should be found with the non-sibilant fricatives, spectrally higher /θ/ and lower /f/, because their energy distributions are broad and overlap with those of vowels. We report four experiments with adult native American English listeners with no speaking or hearing disorders. Exps. 1a,b test the streaming hypothesis's prediction that a /θ-f/ continuum is not differentiated from a following vowel in categorization; Exp. 2 tests streaming's prediction that listeners are unable to predict the quality of a following vowel from the acoustics of the fricative alone. Exp. 3 tests the perceptual differentiation of /s-ʃ/ from following mid vowels as in Exp. 1. Exp. 4 demonstrates that listeners can predict the quality of an upcoming mid vowel from the acoustics of /s/ and /ʃ/ alone.

2. EXPERIMENT 1

2.1. Introduction

Exps. 1a,b tested the generality of the finding that quality of a following vowel influences listeners' categorization of a fricative continuum.

2.2. Method

2.2.1. Stimuli

Two 20-step /θ–f/ continua were made by mixing in complementary proportions 100-ms intervals of fricative noise from /θ/ and /f/ recorded either before /e/ or /o/. In Exp. 1a, the continuum from the /e/ context was spliced onto /eg/ rimes, those from the /o/ context onto /og/ rimes. The vowels' F2-F4 onset frequencies varied incrementally from relatively high to low values in concert with the high-to-low increments in the noise. In Exp. 1b, the final /g/ was replaced with a final /k/ in order to remove a potential lexical bias explanation for the results, because some listeners reported hearing the word *fig*. Final /k/ eliminates this vowel-specific bias, because both *fake* and *folk* are words.

2.2.2. Procedure

Listeners were first trained with correct-answer feedback to respond "f" or "th" as fast as they could to displays of these letters with keywords using 3 randomized trials each for steps 1, 3, 18, and 20. They were then tested without feedback with 28 ran-

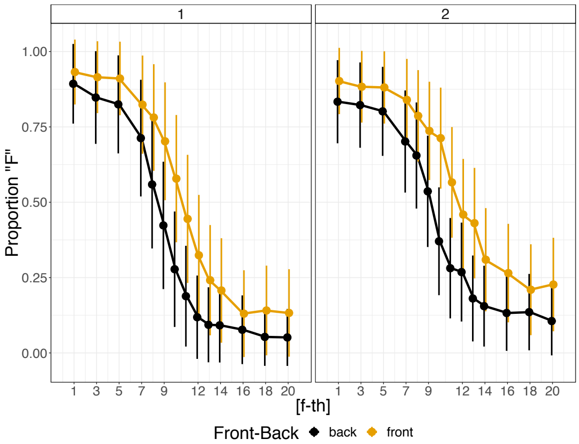

Figure 1: Exps. 1a and 1b: Mean "f" proportions (95% CI)

domized trials each for steps 1, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, and 20 in each vowel context. Each trial comprised a 500 ms display of a cross, the stimulus, a 1500 ms response interval, a 750 ms display of correct-answer feedback on training trials, and a 750 ms inter-trial interval.

2.3. Results

Listeners responded "f" much more before /e/ than /o/ (Fig. 1). A mixed-effects logistic regression model was fitted to "f" responses. Fixed effects were continuum Step, Vowel (e = 0.5, o = -0.5), and Experiment (treatment coded). Step and Vowel were scaled. Decorrelated slopes and intercepts were included for all effects. Table 1 shows that listeners were biased against responding "f" (negative intercept). They responded "f" less often as the fricative became more /θ/-like (negative Step), but more often when the vowel was /e/ than /o/ (positive Vowel) and when the vowel was /e/ than /o/ as Step increased (positive Vowel by Step interaction). They also responded "f" more often and in Exp. 1b than 1a (positive Experiment), but no interactions with Experiment were significant.

Table 1: Exps. 1a,b: Fixed effects estimates

	Estimate	*se*	*z*	*p*
Intercept	-1.168	0.279	-4.180	< 0.001
Step	-2.321	0.196	-11.828	< 0.001
Vowel	0.615	0.092	6.695	< 0.001
Experiment	0.580	0.169	3.425	< 0.001
Vowel:Step	0.154	0.048	3.170	< 0.01

2.4. Discussion

Fricatives were differentiated from vowels. While this could have been produced by a lexical bias in Exp. 1a, that is ruled out in Exp. 1b. Because the energy of non-sibilant fricatives is broadly distributed across frequencies, and overlaps with the energy distribution in the following vowels' spectra, these results cannot be attributed to streaming.

3. EXPERIMENT 2

Exp. 2 used a gating task to test the predictability of a following vowel from preceding fricatives.

3.1. Method

3.1.1. Stimuli

Stimuli were constructed by splicing one of three versions of /θ/ or /f/ onto steps 1, 6, 10, 14, and 20 of a 20-step /e-o/ continuum followed by /ɡ/. One version of each fricative had been recorded before /e/ and /o/, while the third was made by mixing the wave forms of the fricatives recorded before /e/ and /o/ in equal proportions. The first two versions differed acoustically in ways that the listener could use to predict the following vowel's quality, while the third neutralized those differences. Listeners were presented with just the fricative interval (gate 110), the fricative interval plus the first 50 ms of the following vowel (gate 160), which spanned the transition to the vowel's steady-state, or the entire syllable (gate 567). A 500 Hz square wave followed the end of gate and lasted to the end of the stimulus [8], which was 567 ms for all three gates.

3.1.2. Procedure

In the first and fourth trial blocks, listeners were trained with 12 trials with correct-answer feedback to label the vowel as "e" or "o" as fast as they could. These training trials were followed by 120 testing trials using the entire stimulus (gate 567). In the second and third blocks, 24 training trials were followed by 240 test trials using the fricative-alone and fricative+transitions gates (110 and 160). In training trials, all combinations of each of the three versions of the two fricatives with the /e/ and /o/ endpoints of the vowel continuum were presented once for each gate, and in the testing trials, all combinations with the five vowel steps were presented four times. Stimulus order was randomized in both training and testing in all four blocks. Trial timing, response display, and collection were otherwise the same as in Exps.

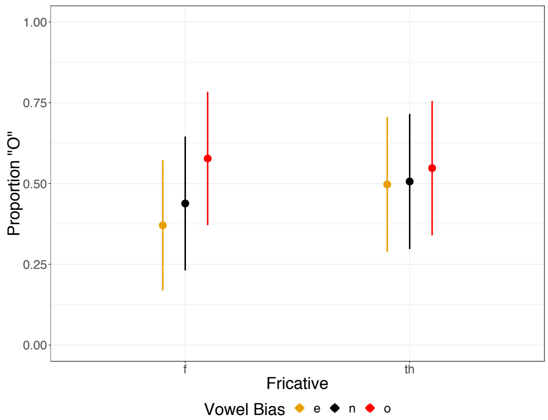

Figure 2: Exp. 2, Gate 110: Mean "o" proportions (95% CI).

1a,b.

3.2. Results

Only responses to the fricative-alone gate (110) tested the prediction of the streaming hypothesis. Listeners responded "o" more to pre-/o/ fricatives, and less to pre-/e/ ones compared to neutralized fricatives (Fig. 2). Both differences are more pronounced with /f/ than /θ/. A mixed effects logistic regression was fitted to "o" responses. Fixed effects were Fricative (/θ/ = 0.5, /f/ = -0.5), Vowel Bias (o = 0.5, no = 0, e = -0.5), and their interaction. Decorrelated intercepts and slopes by participant were included for all effects. Table 2 shows that listeners were not biased to respond "o" or "e" (non-significant intercept) nor did the fricative itself influence "o" responses (non-significant Fricative), but the probability of "o" decreased from the fricative before /o/ to the neutralized fricative to the one before /e/ (positive Vowel Bias). Vowel Bias was reduced for /θ/ compared to /f/ (negative Fricative by Vowel Bias interaction).

Table 2: Exp. 2, Gate 110: Fixed effects estimates

	Estimate	*se*	*z*	*p*
Intercept	-0.009	0.267	-0.032	0.974
Fricative	0.150	0.169	0.889	0.374
Vowel Bias	0.202	0.058	3.499	< 0.001
Fric:V Bias	-0.153	0.042	-3.682	< 0.001

3.3. Discussion

Exp. 2 shows that listeners can also predict the quality of the following vowel from acoustic evidence in the fricative alone, and that they do so better for /f/ than /θ/. Both results undermine the streaming hypothesis because the spectral energy concentration in non-sibilant fricatives overlaps with that in the following vowels, and /f/'s lower frequency concentration overlaps more with vowels' concentration than /θ/'s higher frequency concentration.

4. EXPERIMENT 3

Exp. 3 partially replicated [2, 3, 4, 5, 6, 7, 8, 9, 10]'s finding that listeners categorize more of an /s/ to /ʃ/ continuum as "s" before back vowels than front vowels, by testing listeners before mid /e/ and /o/.

4.1. Method

4.1.1. Stimuli

Fricatives lasted 135 ms. Three 20-step /s-ʃ/ continua were made for Exp. 3. The first two were made in the same way as those continua for Exps. 1a and 1b, and contained vowel-specific information in the fricatives. The third continuum neutralized vowel-specific coarticulatory information by blending endpoints in the same way as the neutralized condition of Exp. 2.

4.2. Procedure

Stimulus presentation was blocked and counterbalanced by vowel-specific versus neutral fricatives. The procedure was the same as in Exps. 1a and 1b, except that within each block, listeners heard 8 training trials, before responding to 20 test trials each for steps 3, 5, 7, 9, 10, 11, 12, 14, 16, 18 in both vowel contexts.

4.2.1. Results

The presence versus absence of coarticulatory information in the fricatives determined the vowel's influence on "s" responses (Fig. 3). When the continuum contained pre-/o/ information, listeners responded "s" more often before /o/ and less often when the continuum did not contain vowel-specific information. Fixed effects were the same as for Exp. 1, plus Fricative (neutral = 0.5, vowel-specific = -0.5) and its interaction with Vowel. Table 3 shows that listeners were biased toward "s" (positive intercept). They responded "s" less often as the fricative became more /ʃ/-like (negative Step), but more often when the fricative was neutral than vowel-specific (negative Fricative), and more often before /o/ when the fricative was vowel-specific, but not when the fricative was neutral (positive Fricative by Vowel in-

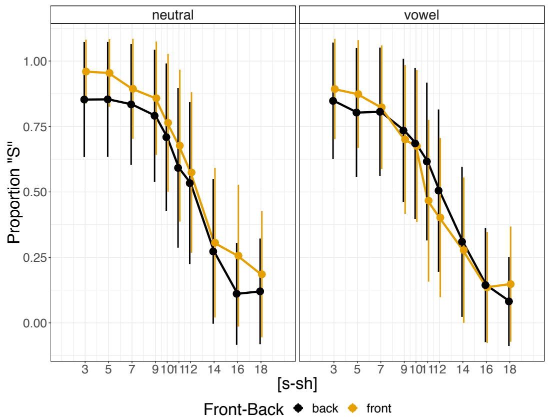

Figure 3: Exp. 3: Mean "s" proportions (95% CI)

teraction).

Table 3: Exp. 3: Fixed effects estimates

	Estimate	*se*	*z*	*p*
Intercept	0.836	0.386	2.163	0.031
Step	-2.339	0.489	-4.784	< 0.001
Fricative	-0.216	0.084	-2.574	< 0.05
Vowel Bias	-0.032	0.158	-0.201	0.840
Fric:V Bias	0.190	0.057	3.320	< 0.001

5. EXPERIMENT 4

Exp. 4 partially replicated [5]'s findings that listeners can predict the quality of an upcoming vowel on the basis of the /s/ or /ʃ/ fricative that precedes it.

5.1. Method

5.1.1. Stimuli

Stimuli were constructed in the same way as for Exp. 2, except with /s/ and /ʃ/, and three gates of just the fricative interval (gate 135), the fricative+transitions (gate 185), and the whole stimulus (gate 582).

5.1.2. Procedure

The procedure was identical to that in Exp. 2.

5.2. Results

Listeners responded "o" most to pre-/o/ fricatives, less to neutralized fricatives, and least to pre-/e/ fricatives (Fig. 4). A mixed effects logistic regression with fixed effects of fricative Bias (/o/ = 0.5, no = 0, /e/ = -0.5), Fricative (/ʃ/ = -0.5, /s/ = 0.5), their interaction, and decorrelated random slopes and in-

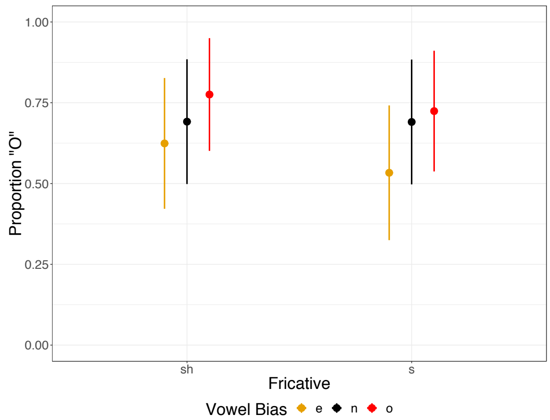

Figure 4: Exp. 4, Gate 135: Mean "o" proportions (95% CI)

tercepts for participants, was fitted to listeners' responses at the 135 ms gate. Table 4 shows that listeners responded "o" most after fricatives from before /o/, less after the no-biased fricatives, and least after fricatives from before /e/ (positive Vowel Bias). Listeners also responded "o" more after /ʃ/ than an /s/ (negative Fricative). Bias and Fricative did not interact.

Table 4: Exp. 4 gate 135: Fixed effects estimates

	Estimate	*se*	*z*	*p*
Intercept	0.82	0.11	7.54	< 0.001
Vowel Bias	0.84	0.11	8.00	< 0.001
Fricative	-0.25	0.09	-2.70	< 0.01

5.3. Discussion

Exp. 4 showed that listeners can anticipate the backness of a mid vowel from a sibilant that contains some of its formant information.

6. GENERAL DISCUSSION

Exps. 1 and 2 showed that non-sibilant fricatives /θ/ and /f/ differentiate from following vowels and and predict those vowels' backness as well as sibilant /s/ and /ʃ/ do. Exps. 3 and 4 replicate earlier findings of differentiation and prediction with sibilants and mid vowels. These parallels between non-sibilants and sibilants undermine any account of differentiation and prediction based in auditory streaming, because non-sibilants' energy is far less concentrated, and far more overlapping with vowels' energy, than sibilants'.

7. REFERENCES

[1] Galle, M., Klein-Pickard, J., Schreiber, K., McMurray, B. in preparation. What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory. *https://doi.org/10.31234/osf.io/5cwvx*.

[2] Mann, V. A., Repp, B. H. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28.3, 213–228.

[3] Mann, V. A., Soli, S. D. 1991. Perceptual order and the effect of vocalic context on fricative perception. *Perception and Psychophysics* 49.5, 399–411.

[4] Mitterer, H. 2006. On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics* 68, 1227–1240.

[5] Nittrouer, S., Whalen, D. H. 1989. The perceptual effects of child-adult differences in fricative-vowel coarticulation. *Journal of the Acoustical Society of America* 86, 1266–1276.

[6] Smits, R. 2001. Evidence for hierarchical organization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance* 27, 1145–1162.

[7] Smits, R. 2001. Hierarchical organization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics* 63, 1109–1139.

[8] Smits, R., Warner, N., McQueen, J. M., Cutler, A. 2003. Unfolding of phonetic information over time: A database of dutch diphone perception. *Journal of the Acoustical Society of America* 113, 563–574.

[9] Whalen, D. H. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of Acoustical Society of America* 69, 275–281.

[10] Winn, M., Rhone, A., Chatterjee, M., Idsardi, W. 2013. The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Frontiers in Psychology* 4, 824.