INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Acoustic properties of strident fricatives at the edges: implications for consonant discrimination

Louis-Marie Lorin[1,*], *Lorenzo Maselli*[2,*], *Léo Varnet*[1,3], *Maria Giavazzi*[1,^]

[1]Département d'études cognitives, École normale supérieure, PSL University, Paris, France
[2]Scuola Normale Superiore, Pisa, Italy
[3]Laboratoire des systèmes perceptifs, CNRS UMR 8248, École normale supérieure, PSL University, Paris, France

louis-marie.lorin@ens.fr, lorenzo.maselli@sns.it, {leo.varnet, maria.giavazzi}@ens.psl.eu

Abstract

Languages tend to license segmental contrasts where they are maximally perceptible, i.e. where more perceptual cues to the contrast are available. For strident fricatives, the most salient cues to the presence of voicing are low-frequency energy concentrations and fricative duration, as voiced fricatives are systematically shorter than voiceless ones. Cross-linguistically, the voicing contrast is more frequently realized word-initially than word-finally, as for obstruents. We investigate the phonetic underpinnings of this asymmetric behavior at the word edges, focusing on the availability of durational cues to the contrast in the two positions. To assess segmental duration, listeners rely on temporal markers, i.e. jumps in acoustic energy which demarcate segmental boundaries, thereby facilitating duration discrimination. We conducted an acoustic analysis of word-initial and word-final strident fricatives in American English. We found that temporal markers are sharper at the left edge of word-initial fricatives than at the right edge of word-final fricatives, in terms of absolute value of the intensity slope, in the high-frequency region. These findings allow us to make predictions about the availability of durational cues to the voicing contrast in the two positions.

Index Terms: temporal marker, duration discrimination, strident fricatives, voicing contrast, perceptual cue

1. Introduction

Speech perception plays an important role in shaping phonological regularities. This was first observed in seminal papers within the theory of Adaptive Dispersion [1] and in studies showing that languages prefer contrastive sound pairs which are perceptually distinct ([2], [3]). The current paper offers a case study within this tradition, focusing on the acoustic underpinnings of the voicing contrast in alveolar strident fricatives.

The contrast between /s/ and /z/ is cross-linguistically more often realized in word-initial than in word-final position (e.g. Russian [4], Polish [5], Standard Dutch [6]). This asymmetry is similar to the one observed in stops for the same featural contrast [7]. In stops, it can be attributed to the difference in the availability of an important perceptual cue to the contrast, i.e. VOT ([8], [9]), as VOT cues at the word edge are only present word-initially. However, this is not a viable explanation for the difference in distribution of the voicing contrast in fricatives. On the other hand, the greater presence of phonetic devoicing in word-final position as compared to word-initially plays a role in shaping the distribution of the voicing contrast [6].

Several studies have described the acoustic properties of voiced and voiceless strident fricatives across different languages (e.g. English [10], [11], [12], European Portuguese [13], [14]). Among the most salient characteristics are frication duration, longer in voiceless fricatives, and low-frequency energy, which is only present in /z/. Furthermore, the duration of a preceding vowel, when present, is shorter for /s/ ([15], [16]). Perceptual results are in general consistent with the observations above, although different acoustic attributes are weighed differently by listeners, with frication duration and the presence of low-frequency energy concentrations being the most salient perceptual cues to discriminate voiced from voiceless fricatives ([10], [17], [18], [19]; cf. [9]). Building on this work, we explore the phonetic grounding of the asymmetry in the realization of the contrast at the word edges, focusing on the availability of durational cues to the contrast in word-initial and word-final position.

Concerning the discrimination of segmental duration, previous research has shown that listeners use amplitude changes as cues to mark segmental boundaries. Kato et al. [20] investigate the perception of temporal modifications in speech by testing the effect of several acoustic characteristics of the temporal markers on listeners' sensitivity to duration modifications. The following conditions appear to be preferred: (i) larger loudness jumps; (ii) sharper slopes; (iii) rising slopes; and (iv) left-edge markers. These results offer a key to understand the distribution of durational contrasts in phonological grammars.

Kawahara et al. show that languages avoid making singleton-geminate contrasts in sonorants (e.g. Ngura, Selayarese, Ilokano, Japanense) because sonorant geminates are easily confused with their corresponding singletons ([21], [22], [23]). They show that the increased confusability of singleton and geminate sonorants as opposed to obstruents is due to the fact that sonorant boundaries involve less amplitude changes ([24], [25]). The blurriness of the segmental boundaries reduces the perceptibility of consonantal duration, which in turn minimizes the cues to the presence of a gemination contrast. The same might apply to other phonological contexts where amplitude changes are involved in boundary detection, namely those where durational cues are hard to weigh due to the lack or paucity of definite markers at the edges.

This paper investigates the acoustic properties of the strident fricatives /s/ and /z/ at the word edges. It tests whether the asymmetry in the distribution of the voicing contrast is (at least partly) due to a similar asymmetry in the distribution of durational cues in the two relevant positions (word-initially

[*] co-first authors, [^] corresponding author

http://dx.doi.org/10.21437/Interspeech.2020-2913

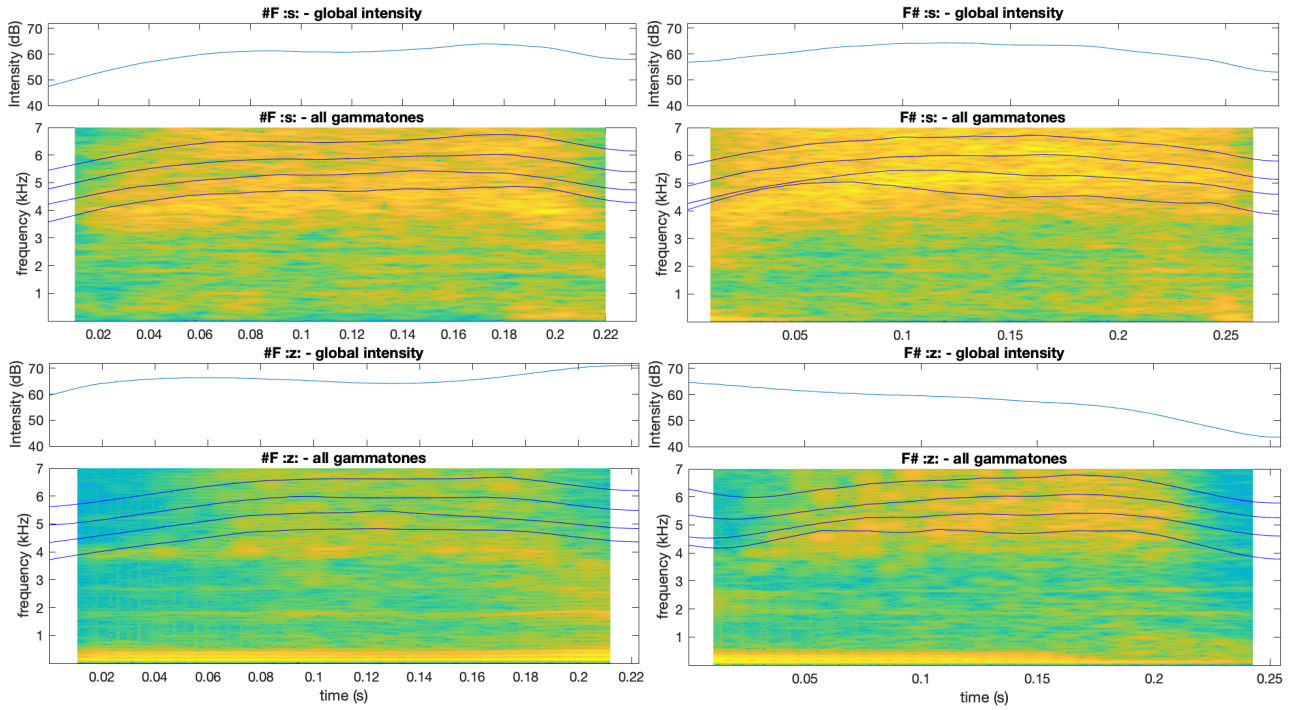

Figure 1: *Global intensity (top) in dB and spectrogram representation superimposed with four gammatones (bottom) from 4 nonce-words tokens ([soof], [voos], [vooz] and [zoof]) (clockwise). Yellower zones correspond to higher energy concentrations. Gammatones variations along the y-axis are in dB.*

and word-finally). Our broad aim is to provide a phonetically grounded explanation of the cross-linguistic distribution of the contrast, but for this study we only use naturally produced American English stimuli as a case in point. We analyze the characteristics of temporal markers at the left edge of word-initial fricatives and those at the right edge of word-final fricatives. Specifically, we analyze (i) the intensity contour in terms of its peak and slope (in absolute value), then (ii) we restrict the analysis to the perceptually relevant high-frequency regions. Results from this phonetic study will allow us to make predictions about the perceptual salience of durational differences at word edges, which will be the subject of a subsequent study.

2. Experiment

2.1. Speakers

Six participants took part in the study (3F: S2-4-6, 3M: S1-3-5), with no observable speech or hearing impairments. All of them are native speakers of American English (mean age = 29.6, SD = $\pm$ 6.4). Speaker F2 (S4) has been excluded from the analysis as she is bilingual. All speakers gave written informed consent.

2.2. Materials and Methods

We selected a list of 12 English words and a list of 12 nonce words compatible with English phonotactics (full list: https://osf.io/rkmfx/?view_only=2c8211efda3e4ac2852ea5edcfbe6751). This material contained alveolar strident fricatives in word-initial, word-final or intervocalic word-medial position (#_V, V_V, V_#). 6 items for each list displayed the English voiceless fricative (/s/) and 6 of them the voiced one (/z/). All tokens are either monosyllabic (for word-initial and word-final fricatives) or disyllabic (for the intervocalic ones). In all tokens, fricatives were adjacent to either low (#_a, a_#, a_a) or high (#_u, u_#, u_i) vowels. 24 filler items were added to each list, resulting in 2 lists of 36 items each, one for words and one for nonce words.

Each speaker read each of the 2 lists 4 times. Items were randomized for each speaker and each repetition and presented in isolation.

2.3. Recordings

Speakers were recorded in a single session in a double-walled sound-proof booth on a TASCAM DR-100MKIII Linear PCM Recorder. Sound files were recorded at a sampling rate of 44.1 kHz and imported into PRAAT 6.1.05 and MATLAB R2020a.

2.4. Analysis

2.4.1. Annotation

Target words and nonce words were segmented in PRAAT. Segmental boundaries for strident fricatives were identified by relying on spectral (energy concentrations) and waveform (lower periodicity) features [26].

2.4.2. Extracting intensity

The stimuli's intensity was analyzed using MATLAB, after extracting all word-final and word-initial fricatives as separate .aiff files. Sound samples were first squared, then low-pass filtered using a second-order butterworth filter (10-Hz cutoff). The shape of the contour was summarized with 3 characteristics: its peak value (in dB) and position, as well as the slope (in dB/s) from the sound onset to the peak (positive slope), and from the peak to the sound offset (negative slope). Slopes were estimated

through linear least square regression.

Another important cue for detecting the fricatives' endpoints is spectral change. Here, we passed the stimuli through a bank of gammatone filters mimicking the frequency resolution of the human ear [27]. Then, we extracted and characterized the intensity contour at the output of 4 gammatones (center frequencies = 4550, 5095, 5700, 6380 Hz); cf. Figure 1. The 4 gammatones where chosen due to their position in the high-frequency region of the spectrum where alveolar strident fricatives exhibit significant spectral power ([28], [29], [30]).

2.4.3. Statistical Analysis

Fillers and items containing fricatives in inter-vocalic position were not included in the analysis, leaving only word and nonce word targets with fricatives in initial and final position. All data were analyzed in RStudio 1.3.1056 using linear mixed-effects models with the lme4 R package [31]. For all models, we included by-participant and by-item random slopes and intercepts and kept the maximal converging random effect structure [32]. All fixed factors were defined using contrast-coding and significance was assessed through comparison of the full model with models without the relevant factor or interaction, with alpha set at 0.05.

A first set of analyses was run to check the durational and voicing characteristics of the target fricatives (/s/ and /z/) at the two edges. The second set of analyses constitutes the focus of the present study, as it investigates the effect of fricative position (word-final vs. word-initial) on two temporal markers: (i) the peak of frication intensity ("Peak") and (ii) the slope's coefficient ("Slope"). As we are interested in the temporal markers at the word edges, we analyze the ascending intensity slope towards the peak for word-initial fricatives, and the descending slope from the peak for word-final fricatives. For this reason, slope analyses were run with the absolute value of the slope coefficient as dependent variable. Analyses of the temporal markers were run both on the whole spectrum and restricting the analysis to the high-frequency region at the 4 gammatones.

3. Results

First, we constructed a model with Duration as the dependent variable and Segment type (/s/ vs /z/) and Position (initial [*#F*] vs final [*F#*]) as fixed factors. The interaction was initially included in the model but it was subsequently removed as it was not significant. The random effect structure included intercepts for participant and item, and a by-participant slope for Position and Segment type. There was a main effect of Segment type, with shorter duration in voiced fricatives than in voiceless ones [β = -0,1, SE = 0.01, $\chi^2(1)$= 20.7, p<0.001], and a main effect of Position, with longer word-final fricatives than word-initial ones [β = 0.1, SE = 0.03, $\chi^2(1)$ = 5.9, p<0.05]. We then constructed a model for voiced fricatives with Devoicing percentage as dependent variable and Position as fixed factor. The interaction was removed as it was not significant. The random effect structure included intercepts for Speaker and Item and an uncorrelated by-participant slope for Position. There was a main effect of Position, with word-final voiced fricatives presenting a greater devoicing percentage than their word-initial counterparts [β = 35.6, SE = 12.9, $\chi^2(1)$ = 5.6, p<0.05].

Next, we focused on the analysis of intensity across the spectrum. First, we constructed a model with Peak as dependent variable and Segment type and Position as fixed factors. The interaction was removed as it was not significant. The random effect structure included intercepts for Speaker and Item, and an uncorrelated by-participant slope for Position. There was a main effect of Position, with lower intensity peaks in word-final position than in word-initial position [β = -2.3, SE = 0.9, $\chi^2(1)$ = 4, p<0.05] and no effect of Segment type; cf. Figure 2.

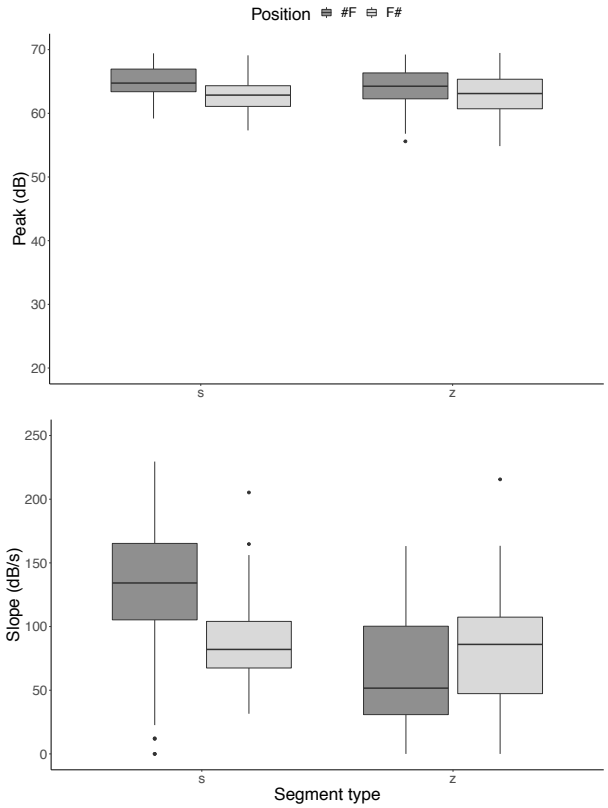

Figure 2: *Global intensity contour: Peak and Slope across the spectrum in word-initial [#F] and w.-final [F#] position, /s/ (left) - /z/ (right).*

Second, we constructed a model with Slope as dependent variable and Segment type (/s/ vs /z/), Position (initial vs final) and their interaction as fixed factors. The random effect structure included intercepts for Speaker and Item and an uncorrelated by-participant slope for Position. There was a main effect of Segment type, with smaller coefficients in voiced fricatives than in voiceless ones [β = -36.5, SE = 6.4, $\chi^2(1)$ = 20.1, p<0.001] and a Position×Segment type interaction [β = 56.5, SE = 12.7, $\chi^2(1)$ = 15, p<0.001], due to the fact that the coefficients were larger in word-initial voiceless fricatives than in word-final ones, but smaller in word-initial voiced fricatives than in word-final ones; cf. Figure 2.

Finally, we restricted the analysis to the high frequency regions. First, we constructed a model with Peak Intensity as dependent variable and Segment type (/s/ vs /z/), Position (initial vs final) and Gammatone (4550 Hz, 5095 Hz, 5700 Hz, 6380 Hz) as fixed factors. The interactions were removed as none of them was significant. The random effect structure included intercepts for Speaker and Item and uncorrelated by-participant slopes for Position and Gammatone. There was a main effect of Segment type, with smaller peaks in voiced fricatives than in voiceless ones [β = -4.0, SE = 1.0 , $\chi^2(1)$ = 11.9, p<0.001] and a main effect of Gammatone, with higher peaks for higher

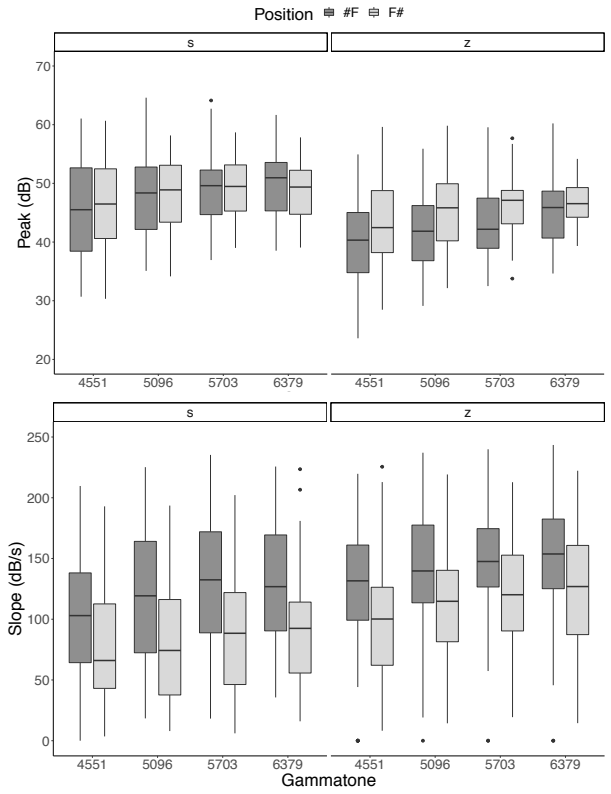

Figure 3: *Gammatone intensity contours: Peak and Slope in word-initial [#F] and w.-final [F#] position, /s/ (left) - /z/ (right).*

gammatones than for the lowest one [β = 0.7, SE = 0.2, $\chi^2(1)$ = 5.6, p<0.05], but no effect of Position; cf. Figure 3.

Second, we constructed a model with Slope as dependent variable and Segment type (/s/ vs /z/), Position (word-initial vs word-final) and Gammatone (4550 Hz, 5095 Hz, 5700 Hz, 6380 Hz) as fixed factors. The interactions were removed as none of them was significant. The random effect structure included intercepts for Speaker and Item and a by-speaker slope for Position. There was a main effect of Segment type, with larger coefficients in voiced fricatives than in voiceless ones [β = 23.5, SE = 6.6, $\chi^2(1)$ = 10.3, p<0.01], a main effect of Position, with smaller coefficients in word-final fricatives than in word-initial ones [β = -32.1, SE = 7.8, $\chi^2(1)$ = 9.7, p<0.01], and a main effect of Gammatone, with a larger coefficient for higher gammatones than for the lowest one [β = 4.5, SE = 0.7, $\chi^2(1)$ = 35.3, p<0.01]; cf. Figure 3.

4. Discussion

After a preliminary analysis of two acoustic properties of the recorded fricatives (duration and voicing), we focused on two temporal markers: (i) the segment's peak intensity and (ii) the absolute value of the slope coefficient at the edge. Analyses were run on the intensity of the whole spectrum and of its higher frequency region.

The first set of analyses confirms that the recorded fricatives are representative of alveolar strident fricatives in American English in terms of durational and voicing properties: (i) voiceless fricatives are longer than voiced ones [10], (ii) word-final fricatives are longer than word-initial ones [33], and (iii) voiced fricatives in final position exhibit devoicing to a higher degree than word-initial ones [34]. The analysis of temporal markers in these fricatives shows that both markers are affected by Position.

With respect to the intensity peak, the whole-spectrum analysis reveals that voiced fricatives have a lower peak than voiceless ones, in line with previous findings [35]. Furthermore, fricatives in word-final position have a lower peak than their initial counterparts. This reflects the production patterns reported in several studies before, as word-initial articulatory strengthening yields increased acoustic intensity ([36], [37]). This effect of Position on the intensity peak was not found when restricting the analysis to the four high-frequency gammatones, suggesting that it is mostly related to low-frequency power sources. A *post-hoc* analysis run separately within each Segment type revealed that there was no effect of Position on the intensity peak in either fricative type.

Our second temporal marker, the slope's coefficient, was also affected by Segment type, but we found no main effect of Position. Interestingly, however, the Position×Segment type interaction in the whole-spectrum analysis reveals that whereas in voiceless fricatives the slope's coefficient is, as expected, greater in word-initial than in word-final position, the reverse holds true for voiced fricatives. This is due to the fact that almost all word-final voiced fricatives are phonetically devoiced to some extent (92.5% of the total number of final [z]'s), which makes them acoustically more similar to voiceless fricatives. This in turn makes their slope's coefficient higher in absolute value; cf. Figure 2. Crucially, when restricting the analysis to the four high-frequency gammatones, phonetic devoicing no longer affects the slope's coefficients. Within the high-frequency region, slope analysis thus reveals that word-initial (ascending) slopes are sharper than their word-final (descending) counterparts for both fricative types; cf. Figure 3.

In sum, we have shown that in word-initial position, strident fricatives at a left edge are characterized by temporal markers which are likely to increase listeners' sensitivity to duration modifications: (i) a greater global intensity peak across the spectrum, and (ii) more salient spectral cues, as estimated by the sharper intensity slopes on the gammatones' intensity contours. Importantly, these spectral cues are not affected by phonetic devoicing in voiced fricatives.

These acoustic results, if confirmed by further perceptual research, would entail some interesting preliminary conclusions as concerns the cross-linguistic preference for voicing contrasts in consonants (and fricatives in particular) to be realized word-initially, i.e. where a left edge is available. As this is by far the word edge where temporal markers are sharper, it is conceivable that this positional privilege plays a role in shaping the phonological contrast (and neutralization thereof) in final position.

Thus, we believe that the present contribution can constitute the ideal first step towards a broader, phonetically-grounded account of the cross-linguistic asymmetry in the distribution of the voicing contrast at the two word edges.

5. Acknowledgements

This study was supported by ANR-17-EURE-0017. We would like to thank Hyesun Cho and Edward Flemming for numerous discussions on this topic in previous phases of this project.

6. References

[1] J. Liljencrants and B. Lindblom, "Numerical simulation of vowel quality systems: The role of perceptual contrast," *Language*, vol. 48, no. 4, pp. 839–862, 1972. [Online]. Available: http://www.jstor.org/stable/411991

[2] E. Flemming, "Auditory representations in phonology. ucla," Ph.D. dissertation, Ph. D. dissertation, 1995.

[3] D. Steriade, "Phonetics in phonology: The case of laryngeal neutralization," 1997.

[4] J. Padgett, "Russian voicing assimilation, final devoicing, and the problem of [v]," *Natural language and linguistic theory*, 2002.

[5] C. Y. Bethin, "Voicing assimilation in polish," *International Journal of Slavic Linguistics and Poetics*, vol. 29, pp. 17–32, 1984.

[6] G. K. Iverson and J. C. Salmons, "Final devoicing and final laryngeal neutralization," *The Blackwell companion to phonology*, pp. 1–22, 2011.

[7] P. Keating, W. Linker, and M. Huffman, "Patterns in allophone distribution for voiced and voiceless stops," *Journal of phonetics*, vol. 11, no. 3, pp. 277–290, 1983.

[8] A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Some cues for the distinction between voiced and voiceless stops in initial position," *Language and speech*, vol. 1, no. 3, pp. 153–167, 1958.

[9] F. Li, A. Trevino, A. Menon, and J. B. Allen, "A psychoacoustic method for studying the necessary and sufficient perceptual cues of american english fricative consonants in noise," *The Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2663–2675, 2012.

[10] K. N. Stevens, S. E. Blumstein, L. Glicksman, M. Burton, and K. Kurowski, "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters," *The Journal of the Acoustical Society of America*, vol. 91, no. 5, pp. 2979–3000, 1992.

[11] M. I. Proctor, C. H. Shadle, and K. Iskarous, "Pharyngeal articulation in the production of voiced and voiceless fricatives," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1507–1518, 2010.

[12] C. H. Shadle, "Acoustics and aerodynamics of fricatives," *The Oxford handbook of laboratory phonology*, pp. 511–526, 2012.

[13] L. M. Jesus and C. H. Shadle, "A parametric study of the spectral characteristics of european portuguese fricatives," *Journal of Phonetics*, vol. 30, no. 3, pp. 437–464, 2002.

[14] C. M. Pinho, L. M. Jesus, and A. Barney, "Weak voicing in fricative production," *Journal of Phonetics*, vol. 40, no. 5, pp. 625–638, 2012.

[15] K. N. Stevens, "Diverse acoustic cues at consonantal landmarks," *Phonetica*, vol. 57, no. 2-4, pp. 139–151, 2000.

[16] D. W. Massaro and M. M. Cohen, "The contribution of fundamental frequency and voice onset time to the/zi/-/si/distinction," *The Journal of the Acoustical Society of America*, vol. 60, no. 3, pp. 704–717, 1976.

[17] A. Jongman, "Duration of frication noise required for identification of english fricatives," *The Journal of the Acoustical Society of America*, vol. 85, no. 4, pp. 1718–1725, 1989.

[18] L. M. Jesus and P. J. Jackson, "Frication and voicing classification," in *International Conference on Computational Processing of the Portuguese Language*. Springer, 2008, pp. 11–20.

[19] H. Cho and M. Giavazzi, "Perception of voicing in fricatives," in *Proceedings of the 18th International Congress of Linguists*, 2009, pp. 986–1006.

[20] H. Kato, M. Tsuzaki, and Y. Sagisaka, "Acceptability for temporal modification of consecutive segments in isolated words," *The Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2311–2322, 1997.

[21] S. Kawahara, "Voicing and geminacy in japanese: An acoustic and perceptual study," *University of Massachusetts occasional papers in linguistics*, vol. 31, pp. 87–120, 2005.

[22] ——, "Sonorancy and geminacy," 2005.

[23] S. Kawahara and M. Pangilinan, "Spectral continuity, amplitude changes, and perception of length contrasts," *The phonetics and phonology of geminate consonants*, vol. 2, p. 13, 2017.

[24] S. Kawahara, "Amplitude drops facilitate categorization and discrimination of length contrasts," *The Institute of Electronics, Information and Communication Engineers, IEICE Technical Report*, vol. 112, no. 145, pp. 67–72, 2012.

[25] S. Kawahara, M. Pangilinan, and K. Garvey, "Spectral continuity and the perception of duration: Implications for phonological patterns of sonorant geminates," *Rutgers University manuscript*, 2011.

[26] K. N. Stevens, *Acoustic Phonetics (Current studies in linguistics; 30)*. MIT Press, 1998.

[27] V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 433–442, 2002.

[28] S. Narayanan and A. Alwan, "Noise source models for fricative consonants," *IEEE transactions on speech and audio processing*, vol. 8, no. 3, pp. 328–344, 2000.

[29] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of english fricatives," *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1252–1263, 2000.

[30] H. Kim, G. N. Clements, and M. Toda, "The feature [strident]," *Features in Phonology and Phonetics: Posthumous Writings by Nick Clements and Coauthors*, vol. 21, pp. 179–194, 2015.

[31] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, G. Grothendieck, P. Green, and M. B. Bolker, "Package 'lme4'," *Convergence*, vol. 12, no. 1, p. 2, 2015.

[32] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of memory and language*, vol. 68, no. 3, pp. 255–278, 2013.

[33] D. H. Klatt, "Linguistic uses of segmental duration in english: Acoustic and perceptual evidence," *The Journal of the Acoustical Society of America*, vol. 59, no. 5, pp. 1208–1221, 1976.

[34] M. Haggard, "The devoicing of voiced fricatives," *Journal of Phonetics*, vol. 6, no. 2, pp. 95–102, 1978.

[35] A. M. Abdelatty Ali, J. Van der Spiegel, and P. Mueller, "Acoustic-phonetic features for the automatic classification of fricatives," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2217–2235, 2001.

[36] O. Fujimura, "Methods and goals of speech production research," *Language and Speech*, vol. 33, no. 3, pp. 195–258, 1990.

[37] C. Fougeron, "Articulatory properties of initial segments in several prosodic constituents in french," *Journal of phonetics*, vol. 29, no. 2, pp. 109–135, 2001.