How prosodic prominence influences fricative spectra in English

Jonathan Barnes[1], *Alejna Brugos*[1], *Stefanie Shattuck-Hufnagel*[2], *Nanette Veilleux*[3]

[1]Boston University [2]Massachusetts Institute of Technology [3]Simmons University
jabarnes@bu.edu, abrugos@bu.edu, sshuf@mit.edu, veilleux@simmons.edu

Abstract

Growing appreciation for non-F0 factors underpinning intonational contrasts has increased attention to interactions between tonal and segmental characteristics of the signal. In some languages, for example, fricatives in certain raised F0 contexts (e.g., final rises) have been shown to bear increased high frequency energy relative to comparable fricatives in low-F0 contexts (Niebuhr's [2009] "segmental intonation"). One approach to this pattern holds that energy peaks during voiceless fricative noise serve as perceptual proxies for the interrupted F0 contour. Speakers actively manipulate fricative spectra to mirror local F0, enhancing tonal contrasts. Results from a study of English fricatives in a variety of metrical and tonal contexts, however, suggest a different explanation: changes in fricative spectral balance are correlated not with local F0, but with intensity of frication noise. Increased "vocal effort" is known to yield both more intense frication noise, and enhanced high-frequency energy, pushing spectral center-of-gravity upward (Shadle & Mair 1996). Subglottal pressure differences under rising and falling pitch in final syllables may affect fricatives similarly (Herman, et al. 1996), supporting fricative spectral balance as a cue not to raised F0, but to increased vocal effort, which, though sometimes correlated with higher F0, does not integrate with it directly.

Index Terms: fricative noise, segmental intonation, perceptual integration

1. Introduction

Recent research increasingly highlights the role of non-F0 factors in the production and perception of contrasting tone and intonation categories. The influence of the segmental composition of the utterance on the perception of pitch contours features prominently here, as numerous studies have focused on the consequences of lowered perceptual salience of F0 during less sonorous segments (e.g., laterals, nasals, voiced obstruents), or over regions characterized by spectral instability (Zhang 2001, Gordon 1999, House 1990). One such line of research (Barnes et al. 2011, 2014) demonstrates that F0 realized over such lower sonority intervals contributes less to judgments of pitch accent timing and scaling than does analogous F0 during, for example, pitch-accented vowels. This has been modelled through the assignment of lower "weights" to low sonority F0 samples during an averaging procedure identified with pitch target scaling perception.

The question of how voiceless obstruents, however, influence perception of the F0 contour has been a matter of some debate. One school of thought observes that, despite such interruptions, F0 contours are "subjectively continuous", suggesting a form of perceptual completion during processing of voiceless segments, such that missing F0 is literally "filled in" to the perceptual record, either through interpolation or extrapolation from existing F0 trajectories. At least for voiceless stops, however, this seems not to occur: Barnes et al. (2011) demonstrate that perceived scaling by English speakers of accented syllables containing voiceless stops is identical to that of shorter syllables with the same realized F0 pattern, but without the gap. Voiceless stop closures in these cases are simply ignored for purposes of tone scaling perception. Mixdorff and Niebuhr (2014) report similar findings in a study focused on prominence perception. In that study, however, voiceless fricatives behaved somewhat differently, in a manner that suggested to them connect with the phenomenon, termed "segmental intonation" by Niebuhr (2009) whereby in several languages, systematically higher fricative spectral centers-of-gravity have been observed in contexts where F0 is also raised. This has been shown most consistently in the case of phrase-final fricatives under rising and falling intonation patterns.

Alongside its systematicity in production (Niebuhr 2012, Ritter & Roettger 2014), segmental intonation has been shown to be available to listeners in perception as well (Niebuhr 2017). One interpretation of the segmental intonation data has involved a form of perceptual integration of the shape and balance of high-frequency fricative noise with the lower frequency patterns in surrounding F0, creating an abstract representation of the pitch contour through regions in which F0 is absent. Speakers might thus manipulate fricative noise as an enhancing cue for tone category identification.

At the same time, however, several studies have noticed that fricative noise intensity covaries in these contexts with spectral center-of-gravity. Niebuhr (2012), in fact, cites studies suggesting that greater acoustic energy in higher frequency bands raises the perceived "pitch" of fricative noise, and therefore that increased intensity may be a vehicle for the transmission of segmental intonation. The relationship between acoustic energy levels and F0 is multifaceted, though, leading us to wonder whether the causality in the case of segmental intonation may not run in a somewhat different direction, one not requiring the perceptual integration of fricative noise and F0. In particular, we hypothesize that contextual variation in subglottal pressure lie at the root both of these intensity patterns, and also of the spectral balance patterns under investigation.

Subglottal pressure changes relate directly to levels of acoustic energy in the signal, whether by increasing the amplitude of vocal fold vibrations in voiced segments, or by increasing the pressure differential across the constriction of a voiceless obstruent. The elevated intensity levels associated with prosodic prominence, in, for example, pitch-accented syllables in canonical intonation languages, are one example of this kind of variation. Those raised intensity levels, furthermore, are not dependent on the tonal composition of the pitch accents in question (Kochanski et al. 2005). Prominent syllables in English are expected to show similar increased intensity, whether bearing Low pitch accents, or High ones.

By the same token though, raised subglottal pressure, along with increased vocal fold tension, is also commonly identified as a means for realizing elevated F0. Indeed, Ladefoged (1967) hypothesizes that speakers may differ systematically in which of these two methods of F0 control they favor, and in which contexts. A long, and still controversial literature (beginning, we believe, with Lieberman 1966) has sought to derive various forms of F0 "downtrends" (including declination, and final lowering) from declining levels of subglottal pressure over the course of the utterance. Whether or not this connection is robustly causal in all such cases, it has been demonstrated clearly (e.g., by Herman et al. 1996) that subglottal pressure drops off rapidly in phrase-final position in English sentences with Low boundary tones, and either fails to fall, or even rises, in English sentences ending with High tonal targets. Importantly, these differences in subglottal pressure, and their corresponding effects on acoustic energy levels in phrase-final syllables, are present regardless of the type of segment that ends the phrase. (That is, they are equally observed in obstruent and sonorant-final utterances.)

At the same time, the increased airflow through a fricative constriction, as would result from raised subglottal pressure, is known to increase the relative concentration of acoustic energy particularly in the higher frequencies of the fricative spectrum, in a manner that is reflected in elevated spectral center of gravity. This is apparently the case, furthermore, without any adjustment to the shape of the constriction or adjacent cavities themselves (Shadle and Mair 1996, Ohala and Solé 2010, Koenig et al. 2013). Increases in "vocal effort" in particular have been demonstrated to produce increases both in fricative intensity, and in spectral balance, yielding fricatives that are perceptually both louder, and higher in "sibilant pitch" (Traunmüller 1987). In principle, therefore, it is possible that higher CoG in certain raised F0 contexts does not involve active manipulation on the part of the speaker, but follows instead as an automatic consequence of increased subglottal pressure. This paper presents the results of an experiment designed to disentangle the relative contributions of local intensity and F0 patterns to systematic variation in fricative spectral balance in a variety of distinct prosodic contexts, and to test the hypothesis that local intensity plays a larger role in determining the spectral peak in frication noise than does an effort by the speaker to signal the shape of the F0 contour during a region of voiceless frication.

2. Methods

We conducted a production experiment in which the English voiceless fricative /s/ appeared in sentential frames designed to manipulate local F0 context and prosodic prominence orthogonally. That is, target fricatives appeared in both high F0 and low F0 contexts, both in prominent and in non-prominent positions. This allowed us to examine the relationships among fricative spectral balance, local F0, and frication noise intensity in detail. Nine native speakers of American English (4 female, 4 males, 1 non-binary) were recorded reading target items embedded in frame sentences cuing various prosodic structures. Recordings were made in a sound-attenuated room, using a headworn condenser microphone. Both sentence frame and target word structure were manipulated to cue differences in 1) position of the target item relative to sentential focus, 2) intonation contour, by way of speech act type (imperative vs. interrogative), and 3) position of the target fricative relative to the lexically stressed syllable of target items. Focus and speech act were manipulated by having subjects read short dialogues (the A role silently, and the B aloud), as in Table 1. In these examples, the target fricative is the initial /s/ in *soda.* Subjects were trained to produce H* L-L% intonation contours in imperative contexts, and L* H-H% in question contexts, which they did with ease. In focused contexts, target items received the nuclear pitch accent, while in post-focus contexts, they followed it.

Table 1: Examples of stimuli manipulating type and location of pitch accent on target words.

Imperative, focused:	A: Should I say coffee again? B: Say *SODA* again, not coffee!
Imperative, post-focus:	A: Should I write soda again? B: *SAY* soda again, don't write it!
Question, focused:	A: Now say soda again. B: Say *SODA* again?! Not coffee?
Question, post-focus:	A: Now say soda again. B: *SAY* soda again?! Not write it?

Target words represented three lexical stress patterns (Table 2), with the target fricative immediately before the stressed vowel, immediately after it, or between two post-tonic vowels. Surrounding vowel context was also controlled.

Table 2: Target words: lexical stress and segmental contexts.

Lexical Stress	-i	-o	-R
stress	seeker seedy	soda soba	service circle
post-stress	Lucy greasy	peso miso	looser racer
distant post-stress	odyssey fantasy		purchaser servicer

Each target item was repeated in each sentence frame three times, for 192 total utterances per speaker (minus fricativeless controls). Target fricatives were manually segmented using Praat (Boersma and Weenink 2019) by experienced annotators. Tokens were eliminated if non-target prosody was produced. The total number of tokens analyzed was 1166.

Target fricatives were high-pass filtered, eliminating spectral energy below 750 hz. For each fricative, duration was measured from onset to offset of frication noise, and spectral center-of-gravity and mean intensity of frication noise were measured during the middle 50% of the fricative's duration. As a proxy for local F0 context, we took six F0 measurements at 20, 40 and 60 ms before the onset and after the offset of frication noise. The mean of the six Z-score normalized F0 measurements was used to represent what F0 might have been realized, had the contour not been interrupted by voicelessness.

3. Results and Discussion

3.1. F0 measurements

Figure 1 displays 95% confidence intervals for the six F0 measurements taken surrounding each target fricative (3 before, 3 after), in each of the elicited prosodic contexts, and confirms that subjects produced the expected F0 patterns for the desired intonation contours (H* L-L% for imperative, and L* H-H% for interrogative). As a result, the accented, stressed tokens (e.g., s̲óda) are flanked by high F0 in the imperative context, but lower F0 in the interrogative, while fricatives following the accented vowels either immediately (e.g., peso) or more distantly (e.g., fantasy) reflect the lower or higher F0 patterns that would be expected given the corresponding rising or falling intonation contours. When focus is shifted from the target to the preceding word, F0 patterns are altered accordingly. For example, in imperatives, the pre-stress fricatives are now preceded by higher F0, but followed by low, since the fall to the post-nuclear trough is already underway. Later fricatives in the same sentence patterns are already fully ensconced in that low F0 trough. For interrogatives, the pre-stress fricatives are preceded by mid-high F0, and followed by high F0, and all later fricatives are in clear high F0 contexts.

Importantly, this means that if fricative spectral center-of-gravity is manipulated by speakers to facilitate perceptual completion of the interrupted F0 track, as predicted by the segmental intonation hypothesis, we expect to find the lowest spectral centers-of-gravity in pre-stress fricatives preceding the L* bearing vowels of pitch-accented words in interrogative sentences (Say **SODA** again?), and in fricatives in the postnuclear trough of imperative sentences with initial focus (**SAY** odyssey again!). We would predict the highest spectral centers-of-gravity to be found in the pre-stress fricatives in words under focus in imperatives (Say **SODA** again!), as well as in the postnuclear plateaux of interrogatives (**SAY** fantasy again?) The biggest differences in spectral center of gravity would be predicted, e.g., between fricatives in the post-nuclear trough/plateau regions of the imperative and interrogative contours respectively.

3.2. Spectral center-of-gravity measurements

Figure 2a shows 95% confidence intervals for measurements of spectral center-of-gravity for /s/-fricatives realized in those same prosodic contexts, making clear that the F0-based predictions sketched in the preceding section are not borne out for English. The two highest centers-of-gravity are found in the pre-stress fricatives in focused imperatives and interrogatives. Indeed, the 95% confidence intervals make these two seem virtually identical, whereas they should be maximally different, since one precedes an H* and the other an L*. Similarly, there is little difference between the accented post-stress and distant post-stress fricatives in the imperative and interrogative sentences, while local F0 suggests that they should differ maximally. Instead, two patterns can be discerned in this data: First, holding all else constant (i.e. speech act and position relative to lexical stress), higher spectral center-of-gravity is observed when the target word is accented than when it is unaccented. That is, prosodic prominence centered on the target word appears to raise spectral CoG, regardless of the local tonal characteristics. Similarly, for each intonational context (i.e. accented-interrogative, unaccented-imperative, etc.), the highest CoG measures are found in the pre-stress fricatives (i.e. those in the onset of the accented syllable). Again, this is true regardless of tonal context. Both these impressionistic observations, furthermore, are supported by the results of a linear mixed regression analysis using vowel quality, focus, stress (encoded binarily as pre-stress vs. other) and speech act as fixed factors, with random intercepts included for target item

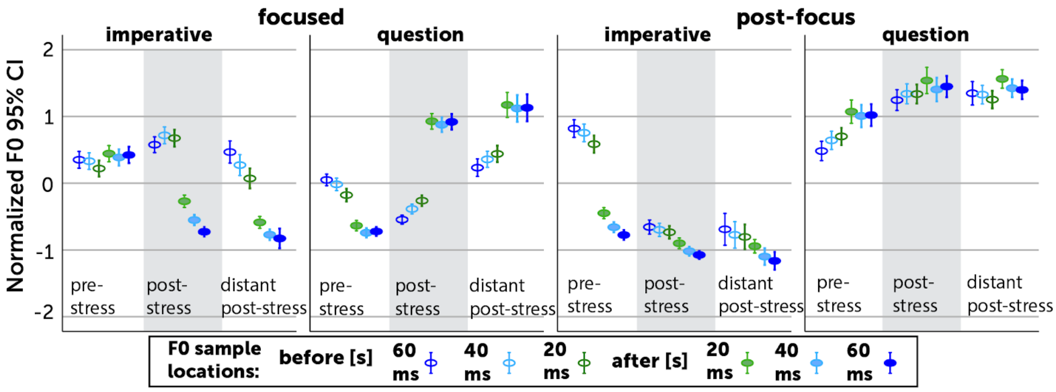

Figure 1: *Normalized F0 patterns in the vicinity of target fricatives for elicited combinations of lexical stress pattern, sentential focus, and speech act type.*

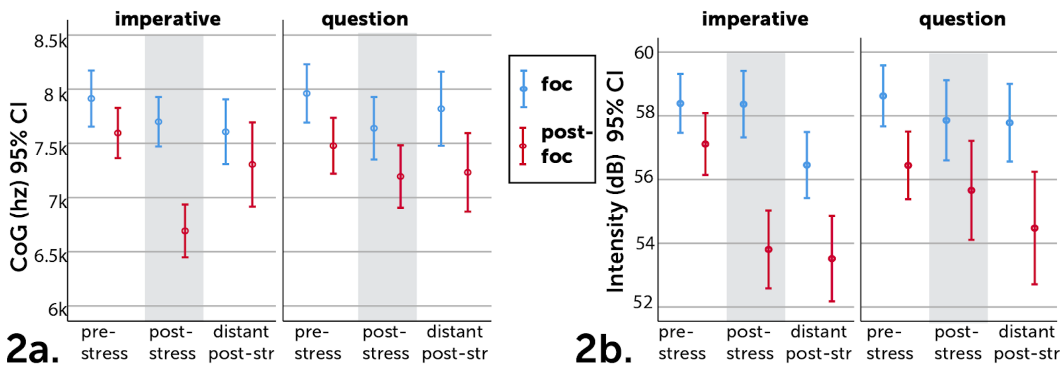

Figure 2: *Center of Gravity (2a) and Intensity (2b) of target fricatives for elicited combinations of lexical stress pattern, sentential focus, and speech act type*

and speaker, conducted with the lme4 package in R (Bates, Maechler & Bolker 2012). An interaction term between stress and question/imperative was included as well. More complex models, including more interactions, and random slopes for speaker and word, failed to converge.

Significance was tested through a series of likelihood ratio tests. The addition of focus to a null model using only vowel quality and question status as fixed factors, was highly significant (chi-square (1) = 64.907, p < .001), focused fricatives showing CoG raised by 407.9 hz (*SE* = 49.93, *t* = 8.171) over unfocused (intercept estimate = 8289.6). The addition of position relative to stress as a further random factor yielded additional improvement (chi-square (2) = 10.969, p = .004), with position in the stressed syllable onset yielding an increase of 466.9 hz (*SE* = 192.97, *t* = 2.42) over other fricatives. Both patterns are consistent with a direct link between positional prominence and higher CoG, unmediated by F0. Other patterns, however, such as the apparently much lower CoG of the imperative, unaccented, post-stress fricatives (**SAY** peso again!), defy our current attempts at explanation.

3.3. Intensity measurements

Figure 2b displays measurements for mean intensity during the same middle 50% of the voiceless fricative, for the same prosodic categories as above. Similarities to Figure 2a are immediately apparent, suggesting the possibility of a causal connection between fricative intensity, a quantifiable manifestation of the "vocal effort" discussed by Shadle and Mair (1996), and spectral center-of-gravity. As before, in each pair of focused/post-focus comparisons, regardless of position in the word or question/imperative status, the fricative in a focused item has greater mean intensity than its post-focus analogue. Similarly, with one exception (the post-stress fricative in focused imperatives), all else equal, fricatives in the onsets of lexically stressed syllables were produced with greater mean intensity than their post-stress counterparts.

We tested the relationship between intensity and spectral center-of-gravity, independent of the linguistic categories around which the study was designed, with another linear mixed effects regression, this time predicting spectral center-of-gravity from vowel quality and mean fricative intensity alone, again with random intercepts for speaker and word. As predicted, the addition of mean intensity (and interactions with vowel quality) to a null model using only vowel quality yielded significant improvement: chi-square (3) = 72.814, p < .001. Adding the measured local F0 mean described above as a predictor to this model produced no improvement (chi-square (1) = .3748, p = .54), nor did the addition of F0 information improve the null model, with intensity removed (chi-square (1) = 2.1482, p = .1427).

3.4. General discussion

These 9 speakers of American English show evidence of a strong relationship between positions of prosodic prominence (both lexical stress and intonational pitch accent as a marker of focus location), and higher spectral center-of-gravity for the voiceless fricative /s/, whether the intonational tone patterns in question situate the fricative in high F0 or low F0 environments. That is, accentedness matters, but accent type (e.g., H* vs. L*) appears not to. To the extent that measured fricative intensity patterns mirror those seen for CoG, we hypothesize that both the increased intensity, and the raised CoG, are the result of increased vocal effort in positions of prominence. The effect of vocal effort on spectral balance in fricatives (presumably via increased airflow through an otherwise unchanged constriction) has been recognized for some time, but has not to our knowledge influenced discussions of how prosody may affect the acoustic characteristics of voiceless obstruents.

A remaining question is whether these results generalize beyond English. Jesus and Shadle (2002) found an effect of vocal effort on fricative spectral balance in Portuguese, but this was not reflected in the realization of lexical stress in that language. It is of couse also possible that the correlation between raised local F0 and higher fricative CoG reported for languages such as German simply does not occur in English.

An additional important question involves the relationship between the findings reported here involving phrase-internal fricatives, and those reported in other studies of segmental intonation, most of which have focused on fricatives at phrase boundaries. Niebuhr (2012), for instance, investigated word-final fricatives in pitch-accented, phrase-final German monosyllables, with rising and falling boundary tone sequences. The prominence of the target words in these cases, traditionally construed, is identical. For our account of segmental intonation to generalize to cases such as this, it would have to be that in German, as in English, final falling contours were associated with rapid drop-off in subglottal pressure, of the kind invoked by Herman, et al. (1996). In this case, it is not immediately clear whether the subglottal pressure patterns in final position should be construed as actually producing the observed F0 changes in rising and falling intonation contours, or whether lowered F0 and lowered intensity phrase-finally are distinct, physiologically independent (if perhaps perceptually interacting) cues to phrase-boundary type. Under either scenario, however, fricative spectral CoG patterns could be considered a manifestation of subglottal pressure changes (present in the relevant contexts in any case both with and without fricatives). No perceptual interaction between the high frequency fricative noise and relatively lower frequency F0 would be necessary in this scenario.

4. Conclusions

This study of the effects of prosodic context on the realization of voiceless fricatives in American English has demonstrated a strong relationship between increased fricative noise intensity in positions of prosodic prominence, and the spectral balance of voiceless fricatives. Local F0 context appeared not to affect fricative realization in any way. This is to be expected, given the established finding that increased vocal effort in fricative production enhances the amount of high-frequency energy in fricative spectra, owing to increased volume velocity airflow through the constriction. In view of the lowered subglottal pressures observed in English under conditions of "final lowering", these patterns may also explain the correlation in some positions, in some languages, between higher F0 contexts, and higher fricative centers-of-gravity. If this account does generalize, it is possible that the pattern that has been termed "segmental intonation" does not involve fricatives providing information to listeners about F0 patterns directly, as has been hypothesized. Instead, fricative spectral patterns may serve as cues to prominence and/or vocal effort levels. In some cases, increased vocal effort may coincide with raised F0, and some perceptual interaction between the two, at a higher level, may occur in some contexts. Direct integration of frication noise in F0 contour perception, however, would not need to.

5. References

[1] J. Barnes, A. Brugos, N. Veilleux & S. Shattuck-Hufnagel, “Voiceless Intervals and Perceptual Completion in F0 Contours: Evidence from Scaling Perception in American English,” *International Congress of Phonetic Sciences 17*, Hong Kong, 2011.

[2] J. Barnes, A. Brugos, N. Veilleux & S. Shattuck-Hufnagel, “Segmental Influences on the Perception of Pitch Accent Scaling in English,” In *Proceedings of Speech Prosody 7*, N. Campbell, Gibbon, and D. Hirst (eds.), pp. 1125-1129, 2014.

[3] D. Bates, M. Maechler, B. Bolker & S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, 67(1), 1-48, 2015.

[4] P. Boersma & D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.1.08, retrieved 5 December 2019 from http://www.praat.org/

[5] M. Gordon, *Syllable weight: Phonetics, phonology, typology,* PhD Thesis, UCLA, 1999.

[6] R. Herman, M. Beckman & K. Honda, "Subglottal pressure and final lowering in English", In *Proceedings of the International Conference on Spoken Language Processing,* 145-148, 1996.

[7] D. House, *Tonal perception in speech*. Lund, Sweden: Lund University Press, 1990.

[8] L. Jesus & C. Shadle, “A parametric study of the spectral characteristics of European Portuguese fricatives,” *Journal of Phonetics*. 30. 437-464. 10.1006/jpho.2002.0169, 2002.

[9] L. Koenig, C. Shadle, J. Preston & C. Mooshammer, “Toward improved spectral measures of /s/: Results from adolescents,” *Journal of Speech, Language & Hearing Research* 56(4): 1175–1189. doi:10.1044/1092-4388(2012/12-0038). 2013

[10] G. Kochanski, E. Grabe, J. Coleman, & B. Rosner, “Loudness predicts prominence: Fundamental frequency lends little.” *Journal of the Acoustical Society of America* 118 (2), 1038-1054, 2005

[11] P. Ladefoged, *Three areas of experimental phonetics*, London: Oxford University Press, 1967.

[12] P. Lieberman, *Intonation, perception, and language.* Ph.D. dissertation, Massachusetts Institute of Technology, 1966.

[13] H. Mixdorff, A. Hönemann, O. Niebuhr & C. Draxler, “Perceived Prominence Reflected by Imitations of Words with and without F0 Continuity,” 10.13140/2.1.2990.8487. *7th International Conference on Speech Prosody*, Dublin, Ireland, 2014.

[14] H. Mixdorff, & O. Niebuhr, “The influence of F0 contour continuity on prominence perception,” *14th Annual Conference of the International Speech Communication Association*, Lyon, France, 2013.

[15] O. Niebuhr, "Intonation segments and segmental intonation", In *INTERSPEECH-2009*, 2435-2438, 2009.

[16] O. Niebuhr, “At the edge of intonation – The interplay of utterance-final F0 movements and voiceless fricative sounds,” *Phonetica* 69, 7–27, 2012.

[17] O. Niebuhr, “On the perception of “segmental intonation”: F0 context effects on sibilant identification in German,” *EURASIP Journal of Audio, Speech & Music Processing 19*, 2017.

[18] J. Ohala & M-J Sole, “Turbulence and phonology. Turbulent sounds. An interdisciplinary guide”, in S. Fuchs, M. Toda and M. Zygis (eds.), *Turbulence and phonology*, Berlin: Mouton deGruyter, pp.37-97, 2010.

[19] S. Ritter & T. Roettger, “Speakers modulate noise-induced pitch according to intonational context,” in *Proceedings of the 7th International Conference on Speech Prosody,* Dublin, 2014.

[20] C.H. Shadle & S.J. Mair, “Quantifying spectral characteristics of fricatives.” *Proceedings of the International Conference on Spoken Language Processing* [ICSLP] Philadelphia. pp. 1521–1524, 1996.

[21] H. Traunmüller, “Some aspects of the sound of speech sounds,” In M.E.H. Schouten (ed.), *The Psychophysics of Speech Perception*. Dordrecht: Martinus Nijhoff, 293-305, 1987.

[22] P. Welby & O. Niebuhr, “Segmental intonation information in French fricatives,” *Proceedings of the International Congress of Phonetic Sciences*, Melbourne 2019.

[23] J. Zhang, *The Effects of Duration and Sonority on Contour Tone Distribution— Typological Survey and Formal Analysis*. PhD Thesis, UCLA, 2001.