SEGMENTAL INTONATION INFORMATION IN FRENCH FRICATIVES

Pauline Welby[1] and Oliver Niebuhr[2]

[1]Aix Marseille University, CNRS, LPL, Aix-en-Provence, France
[2]Centre for Industrial Electronics, University of Southern Denmark, Sønderborg, Denmark

pauline.welby@univ-amu.fr, olni@sdu.dk

ABSTRACT

We examined the "segmental intonation" hypothesis [17], according to which voiceless consonants contain spectral information that may contribute to the percept of high or low pitch in the absence of fundamental frequency (F0). French speakers read target words embedded in a carrier phrase and containing fricatives in accentual phrase-initial, -medial or -final position (e.g. *sidéré* 'stunned', *nécessite* 'require', *ressaisisse* 'seize again'), expected to correspond to regions of low, intermediate or high F0, respectively, as well as control words containing only sonorants (e.g. *laminé* 'rolled'). Analyses show lower center of gravity (CoG) for word-initial (low F0 region) than for word-final (high F0 region) fricatives. For word-final fricatives, CoG is higher at the end than in the beginning of the fricative, which may contribute to the percept of the continuation of the F0 rise across the preceding vowel.

Keywords: segmental intonation, F0 perturbation, sibilant pitch, French, accentual phrase, CoG.

1. INTRODUCTION

A small but growing body of research has shown that the spectral characteristics of voiceless consonants vary according to intonational context, in line with the "segmental intonation" hypothesis [17]. Rather than simply perturbing the fundamental frequency curve of an utterance, as has been traditionally assumed [9, 2], voiceless consonants have been shown to contain spectral information that may contribute to the percept of high or low periodic pitch in the absence of fundamental frequency (F0).

Frication noise is known to create pitch impressions in listeners in the absence of F0. For example, in his chapter on fricatives, Johnson [7] notes: "When you make a series of fricatives starting from a pharyngeal constriction and moving the constriction forward to the alveolar ridge, you may be able to hear a change in the 'pitch' of the fricative" (p. 157). These pitch impressions, however, have traditionally only been considered to be relevant in particular contexts, such as psychoacoustic contexts ("sibilant pitch", [27]) or whispered speech [11]. However, studies by Niebuhr [15–19] have shown that speakers vary the spectral energy distributions of fricatives systematically according to the adjacent F0 context. Furthermore, perception experiments have provided evidence that these spectral differences in voiceless obstruents create pitch impressions in listeners that are integrated with the voiced parts of an utterance and interpreted as part of the global intonation pattern [17], similar to the "picket-fence effect" in psychoacoustics [4]. Other studies have refined the notion of segmental intonation [6] and provided evidence for it in languages like Polish and English [24, 31, 32, 21]. These studies have all shown that spectral center of gravity (CoG) can capture segmental intonation with respect to both the acoustic manifestations and their reflections in the perception of fricative pitch.

Most previous studies have examined intonational differences at major prosodic boundaries conveying pragmatic differences such as sentence modality (question vs. statement), e.g. [17, 18]. This has left open the question of whether segmental intonation is produced with the (implicit) goal of perceptually filling in gaps in F0 or reflects a meaning-driven change in articulatory effort. The only studies looking at segmental intonation utterance-medially are both on German [18, 13]. However, to decide between segmental intonation being a filler of F0 gaps or a meaning-driven effort change, cross-linguistic studies examining similar F0 gaps in different contexts are required. One goal of the current study is to help resolve this question.

2. METHODS

We examined whether the CoG of French voiceless fricatives differed according to whether they appeared adjacent to regions of low, intermediate or high F0. CoG was selected because it is closely correlated with perceived fricative pitch and has been found to respond very sensitively to changes in fricative articulation across languages (see [1, 22]), and also because it allows us to compare our results with those of previous studies.

2.1. Materials

To place fricatives in regions expected to be produced with low, high or intermediate F0, we exploited a tonal scaling regularity reported in the literature. In French, accentual phrases (APs) that are not utterance-final are typically realized with an F0 rise that reaches its peak late in the last full syllable of the AP. If an AP is long enough, another rise is often realized with its starting point at the beginning of the first content word of the AP. Typically, APs that contain at least three content word syllables are produced with this two-rise (early rise-late rise) $L_1H_1L_2H_2$ pattern [8, 28]. Crucially, the late rise typically begins at a higher F0 than the early rise, i.e. L_2 is higher than L_1 [28].

The critical intonational differences here do not convey pragmatic differences. They are, however, important for listeners in distinguishing between the early rise (L_1H_1), which is a cue to a content word beginning in word segmentation and lexical access [29, 26], and the late rise (L_2H_2), whose peak signals the end of the AP [12]. Previous research has established differences in the realization of these two rises, including the scaling differences outlined above, as well as differences in the shape of the peaks (simple vs. plateau) and the slope of the rise [25, 29]. Preliminary results suggest that voiceless segments render these intonational cues less salient, particularly when they interrupt F0 turning points [30].

A set of target words was constructed using the Lexique database [14] to identify candidates with the required properties. They were three-syllable content words, with a voiceless fricative in word-initial, -medial or -final position. Given the tonal scaling of the $L_1H_1L_2H_2$ pattern, these positions correspond to relatively low, intermediate or high F0 in the adjacent region. Examples are given in Table 1.

Table 1: Example target words for /s/

Position in $L_1H_1L_2H_2$ AP	Expected relative F0 of adjacent region	Word	Gloss
initial	low	sidéré	‘stunned’
medial	intermediate	nécessite	‘require’
final	high	ressaisisse	‘seize again’
all-sonorant control		laminé	‘rolled’

The three voiceless fricatives of the French phoneme inventory – /s/, /ʃ/, and /f/ – were included in the recordings. The segmental context surrounding the target fricative was controlled to minimize the influence of coarticulation. Four words were selected for each of the three positions, giving a set of 12 items per fricative. In addition, there were six control words containing only sonorant segments.

Each item was repeated three times in the list, and the order of the list was pseudo-randomized. Here we report results for the sibilant fricative /s/. Analysis of the other two fricatives is ongoing. Target words were embedded in the carrier in (1).

(1) Et « X » t'était proposé.
‘And “X” was suggested to you.’

2.2. Participants

Participants were 15 female native speakers of French, who had been raised in mainland France, and were currently attending university.

2.3. Procedures

Participants were instructed that they would read sentences with the structure given in (1), which could be imagined as part of a conversation about a crossword puzzle.

Each participant was recorded individually in a sound attenuated room. The words in the carrier phrase were presented one-by-one on a computer screen. Participants were instructed to read the sentence first silently and then out loud, at a normal speaking rate. Recordings were made using an AKG C520 head-worn microphone and a Zoom H4nSP Handy recorder at a sampling rate of 44.1 kHz.

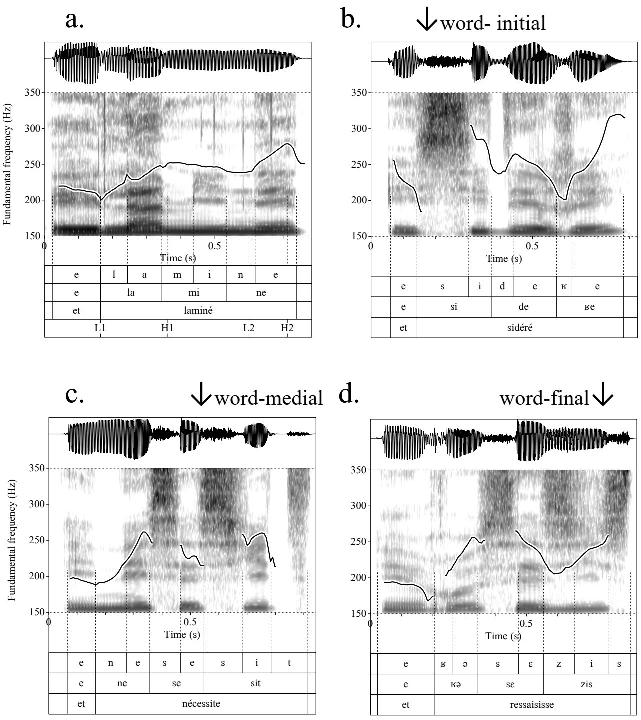

Figure 1. Example productions of a.) an all-sonorant control word and target words with /s/ in b.) word-initial, c.) word-medial, and d.) word-final position, $L_1H_1L_2H_2$ AP pattern.

Each sentence was saved as individual sound file, and an automatic segmental and syllabic annotation was generated using the EasyAlign forced alignment plug-in [5] for Praat [3]. For each sound file, the waveform and spectrogram and associated annotation were inspected, and clear errors in critical boundaries (e.g. beginning and end of target fricatives) were manually corrected. Example productions are shown in Fig. 1.

2.4. Analyses

Acoustic analyses were performed in Praat, and statistical analyses were performed using generalized linear mixed models in R [23] and the package *lme4*.

3. RESULTS

3.1. Intonation patterns

A first step inspection of the all-sonorant control items (e.g. *laminé*) was manually performed to confirm that speakers had produced the required two-rise $L_1H_1L_2H_2$ pattern across the target AP, resulting in the exclusion of the data of six of the 15 participants. These participants produced a variety of other patterns, including falls, e.g. a slightly falling $L_1H_1H_2$, and pitch patterns consistent with narrow focus (see [20] for this common issue with sentence-list elicitations). In a second step automated by a Praat script, the F0 of the control items was measured at three different points: midpoint of the first third of syllable 1, midpoint of the first third of syllable 3, midpoint of the last third of syllable 3, corresponding to word-initial, -medial, and -final position, respectively, for the critical target words. After this second check, the data from one additional participant were excluded. All analyses are based on the data of the eight participants who produced the required intonation pattern (see Fig. 2).

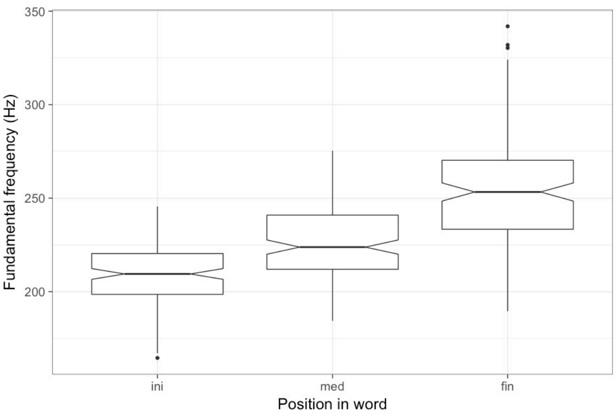

Figure 2. Mean F0 of all-sonorant control items at word-initial, -medial, and -final position.

A linear mixed-effects model was fitted to the mean F0 values, with Position in the word (word-initial, word-medial, word-final) as fixed effect, Participant and Item as random intercepts. A significant difference was found between F0 in word-initial position (209 Hz, SD = 19) and in word-medial position (225 Hz, SD = 21, b = 16.304, SE = 1.821, t =8.954 , p < 0.0001), and between F0 in word-initial position and in word-final position (253 Hz, SD= 34, b = 43.826, SE = 1.821, t = 24.069, p < 0.0001). Note that because there is a rise across the final syllable, F0 in word-final position is higher than that in word-medial position (see Fig. 1a).

We then took CoG measurements within a frequency range of 2 and 12 kHz on the basis of spectral slices, with a 10 ms Gaussian window centered on the first, second, and last third of the fricative. Mean CoG was calculated on the basis of these slice values.

3.2. Center of gravity

We first compared mean CoG across the three positions in the word (Fig. 3).

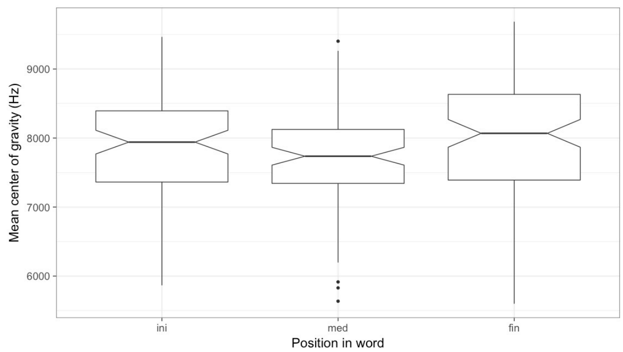

Figure 3. Mean center of gravity for /s/ by position of the fricative in the word.

A linear mixed-effects model was fitted to the CoG values, with Position in the word (word-initial, word-medial, word-final) as fixed effect, Participant and Item as random intercepts. A significant difference was found between the mean CoG in word-initial position (7860 Hz, SD = 995) and in word-final position (7960, SD = 989, b = 141.26, SE = 76.54, t = 2.285, p = 0.0231). This CoG difference of about 100 Hz is slightly smaller than those reported in previous studies (e.g. [17, 21]), but note the same also applies to the associated F0 difference. No difference between mean CoG in word-initial position and in word-medial position was found (7688 Hz, SD = 874, b = -141.26, SE = 76.54, t = -1.846, p = 0.0661).

It is clear from Figs 1 and 2 that the position of the fricative in the word corresponded to the expected differences in adjacent F0 (initial: low F0, final: high F0, medial: intermediate F0). There is, however, another intonational difference between the word-final position and the other two positions, one that is apparent in the all-sonorant control items (Fig. 1a). In the materials, word-initial fricatives are at the boundary between the proclitic function word *et* and a content word and thus straddle a (missing) F0 "turning point", the low starting point of the early rise (L_1H_1) of the $L_1H_1L_2H_2$ pattern. Word-medial fricatives, which are in onset position of the last syllable, also often straddle a missing turning point, the Low starting point of the late rise (L_2H_2). The realization of a word-medial fricative at a turning point is less systematic, however, since L_2 can be realized earlier or later, for example, well into the vowel of the last syllable. The word-final position, however, corresponds to the continuation of a fairly linear rise in F0 starting from the L_2 of the late rise (L_2H_2) and continuing across the vowel nucleus of the last syllable. We examined whether this difference was reflected over the time course of the fricatives.

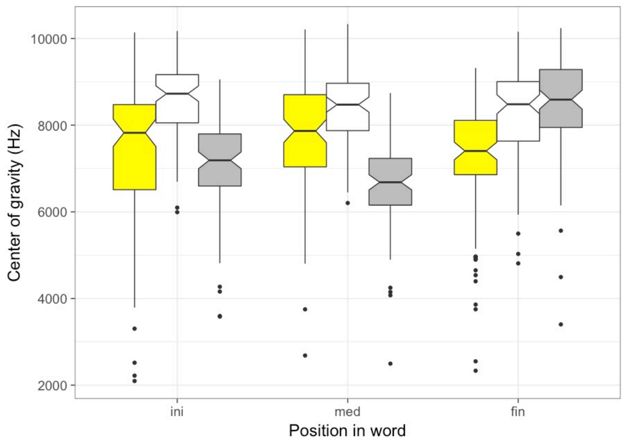

Figure 4. Center of gravity for /s/ by position in word and fricative time slice: 1st third (yellow), 2nd third (white), 3rd third (gray).

Fig. 4 shows CoG across the first, second, and third time slices (thirds) for each of the positions of the fricative in the word. In word-final position, there is a steady CoG rise from the first to the last fricative time slice. A linear mixed-effects model was fitted to the CoG values with Time slice (1st third, 2nd third, 3rd third) as fixed effect, Participant and Item as random intercepts. CoG in the last third (8524 Hz, SD = 1029) was higher than CoG in the second third (8245 Hz, SD = 1091, b = -313.2, SE = 130.954, t =-2.392, p = 0.0174) and from CoG in the first third (7347 Hz, SD = 1164, b = -1283.05, SE = 133.76, t =-9.592 , p < 0.0001). For /s/ in both word-initial and in word-medial position, CoG is higher at the midpoint (second third) than that at the first third and the last third (all p's <0.0001).

3.2. Intensity

A final mixed effects model examined mean fricative intensity, since this parameter can also cause changes in perceived fricative pitch (see [17]). No differences were found. We thus have no evidence that intensity differences play a role in French segmental intonation.

4. DISCUSSION

The results provide confirmation for and extend a number of research findings on segmental intonation. First, sibilants in regions of low F0 have lower mean CoG than those in regions of high F0. This cue to fricative pitch had been found in many other studies for a number of other languages [15–19, 21, 24, 31, 32], but never in non-utterance-final position in a language other than German [13, 19]. CoG differences in most earlier studies were larger, but also co-occurred with larger F0 differences. The present study found a CoG difference of 100 Hz for an F0 difference of about 40 Hz (or 3.2 st). In [17], a CoG difference of about 500 Hz was found for an F0 difference of about 120 Hz (or 12.3 st). Across all studies, a rule of thumb emerges that for /s/ every 2–3 st causes a CoG change of about 100 Hz. Different proportions seem to apply to other fricatives.

Second, unlike in earlier studies, the critical F0 differences were not associated with meaning differences. The current study provides the first clear empirical evidence that segmental intonation does not reflect a meaning-driven change in articulatory effort, but rather is produced with the (implicit) goal of perceptually filling in gaps in F0 [17, 18].

Third, the development of CoG over time differed for fricatives that interrupted fairly linear stretches of F0 (word-final /s/) and those that interrupted turning points (word-initial and -medial /s/). This finds a parallel (albeit a preliminary one) to findings that the voiceless fricatives can perceptually bridge gaps in F0 when they interrupt a fairly linear stretch of F0, but not when the perturbation occurs at an F0 turning point [13]. Similarly, [24] also stressed the need to look at CoG time courses. This study found differences in the development of CoG over time between low and high boundary tones, pointing to tonal-target realizations inside voiceless fricative segments. In our study, the fact that the CoG across word-final /s/ continues the F0 rise on the preceding vowel also points to a "non-tonal" tonal target at the end of the /s/. Other researchers have reported voice quality correlates to pitch targets [10]. These findings suggest that models of intonational phonology should take into account correlates of pitch other than F0.

5. REFERENCES

[1] Al-Khairy, M. A. 2005. *Acoustic characteristics of Arabic fricatives*. PhD thesis, University of Florida.

[2] Barnes, J., Brugos, A., Veilleux, N., Shattuck-Hufnagel, S. 2011. Voiceless intervals and perceptual completion in F0 contours: Evidence from scaling perception in American English. *ICPhS,* Hong Kong, 108–111.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0, retrieved from http://www.praat.org.

[4] Bregman. A. S. 1990. *Auditory Scene Analysis*. Cambridge, MA: M.I.T. Press.

[5] Goldman, J.-P. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. *Interspeech,* Florence.

[6] Heeren, W. F. L. 2015. Coding pitch differences in voiceless fricatives: Whispered relative to normal speech. *Journal of the Acoustical Society of America* 138, 3427–3428.

[7] Johnson, K. 2012. *Acoustic and Auditory Phonetics*. Oxford: Wiley-Blackwell.

[8] Jun, S.-A., Fougeron, C. 2002. Realizations of accentual phrase in French. *Probus* 14, 147–172.

[9] Kohler, K. J. 1990. Macro and micro F0 in the synthesis of intonation. In: Kingston, J., Beckman, M. E. (eds) *Papers in Laboratory Phonology I*. Cambridge University Press: Cambridge, 115–138.

[10] Kuang, J. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America* 142, 1693–1706.

[11] Meyer-Eppler, W. 1957. Realization of prosodic features in whispered speech. *Journal of the Acoustical Society of America* 29, 104–106.

[12] Michelas, A., D'Imperio, M. 2010. Accentual Phrase boundaries and lexical access in French. *Speech Prosody*, Chicago.

[13] Mixdorff, H., Niebuhr, O. 2013. The influence of F0 contour continuity on prominence perception. *Interspeech*, Lyon, 230–234.

[14] New, B. 2006. Lexique 3 : Une nouvelle base de données lexicales. *Traitement Automatique des Langues Naturelles*, Leuven.

[15] Niebuhr, O. 2008. Coding of intonational meanings beyond F0: Evidence from utterance-final /t/ aspiration in German. *Journal of the Acoustical Society of America* 124, 1252–1263.

[16] Niebuhr, O. 2009. Intonation segments and segmental intonations. *Interspeech,* Brighton, 2435–2438.

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

[18] Niebuhr, O. 2017. On the perception of "segmental intonation": F0 context effects on sibilant identification in German. *EURASIP Journal on Audio, Speech, and Music Processing* 19. DOI 10.1186/s13636-017-0115-3.

[19] Niebuhr, O., Lill, C., Neuschulz, J. 2011. At the segment-prosody divide: The interplay of intonation, sibilant pitch and sibilant assimilation. *ICPhS*, Hong Kong, 1478–1481.

[20] Niebuhr, O., Michaud, A. 2015. Speech data acquisition: The underestimated challenge. *KALIPHO* 3, 1– 42.

[21] Percival, M., Bamba, K. 2017. Segmental intonation in tonal and non-tonal languages. *Journal of the Acoustical Society of America* 141, 3701.

[22] Prasad, R. S., Yegnanarayana, B. 2018. Identification and classification of fricatives in speech using zero time windowing method. *Interspeech*, Hyderabad, 187–191.

[23] R Core Team. 2018. R: A language and environment for statistical computing [computer software manual]. Vienna. Retrieved from http://www.r-project.org/ (version 3.5.1).

[24] Ritter, S. Roettger, T. B. 2014. Speakers modulate noise-induced pitch according to intonational context. *Speech Prosody*, Dublin, 890–893.

[25] Rolland, G., Lœvenbruck, H. 2002. Characteristics of the accentual phrase in French: An acoustic, articulatory and perceptual study. *Speech Prosody*, Aix-en-Provence, 611–614.

[26] Spinelli, E., Grimault, N., Meunier, F., Welby, P. 2010. Intonational cue to word segmentation in phonemically identical sequences. *Attention, Perception and Psychophysics* 72, 775–787.

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

[28] Welby, P. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34, 343–371.

[29] Welby, P. 2007. The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication* 49, 28–48.

[30] Welby, P., Niebuhr, O. 2016. The influence of F0 discontinuity on intonational cues to word segmentation: a preliminary investigation. *Speech Prosody,* Boston, 40–44.

[31] Żygis, M., Pape, D. Jesus, L., Jaskula, M. 2014. How do voiceless fricatives contribute to intended intonation? A comparison of whispered, semi-whispered and normal speech. *10th International Seminar on Speech Production*, Cologne, 472–475.

[32] Żygis, M., Pape, D., Koenig, L. L., Jaskuła, M., Jesus, L. M. T. 2017. Segmental cues to intonation of statements and polar questions in whispered, semi-whispered and normal speech modes. *Journal of Phonetics* 63, 53–74.