



The Influence of F0 Discontinuity on Intonational Cues to Word Segmentation: A Preliminary Investigation

Pauline Welby¹, Oliver Niebuhr²

¹ Aix Marseille Université, LPL, CNRS, 5 avenue Pasteur, 13100 Aix-en-Provence, France

² Department of Design and Communication, IRCA, University of Southern Denmark, Denmark

pauline.welby@lpl-aix.fr, olni@sdu.dk

Abstract

The paper presents the results of a 2AFC offline word-identification experiment by [1], reanalyzed to investigate how F0 discontinuities due to voiceless fricatives and voiceless stops affect cues to word segmentation in accentual-phrase-initial rises (APRs) of French relative to a reference condition with liquid and nasal consonants. Although preliminary due to the small sample size, we found initial evidence that voiceless consonants degrade F0 cues to word segmentation relative to liquids and nasals. In addition, this degradation seems to be stronger for voiceless stops than for voiceless fricatives, as listeners in the latter condition were more sensitive to (resynthesized) changes in the residual rise fragments. This evidence is consistent with the intonational model of [2], which predicts that listeners can to some degree restore frication-filled gaps but not silent gaps in F0, by using pitch impressions created by the fricative noise. Our results call for follow-up studies that use French APRs as a testing ground for this intonational model and also examine the precise nature of intonational cues to word segmentation.

Index Terms: word segmentation, segmental perturbation, segmental intonation, F0 rise, French

1. Introduction

We now know that listeners use intonational patterns of their native language to guide word segmentation, the process of locating word boundaries in the speech stream ([3], [4], for French; [5] for Korean; [6] for Japanese *inter alia*), and lexical access, the retrieval of words from the mental lexicon ([1]).

One issue that has not yet been directly addressed, to our knowledge, is the extent to which intonational cues to word segmentation are reliably available to listeners, in particular the effect of a discontinuous F0 curve on these cues. This question was raised (but left open) by [3] who wrote of the generalizability of intonational cues to word segmentation in French:

"...we must ask how general a cue the early [F0] rise and the precise alignment of its starting point to the beginning of a content word [is]. There are at least two possible scenarios — the first is that the cue is mostly useful for sequences of sonorants or sequences of voiced consonants, where there is minimal microprosodic perturbation [...]. The other possibility is that the listener is well-equipped to deal with microprosodic perturbations and that these perturbations do not hinder speech perception. After all, segmentally conditioned f0 perturbation is pervasive in all spoken languages, yet we do not perceive pitch as discontinuous [...].

It seems plausible that listeners [...] can use microprosodic perturbations to help them identify segments, while factoring out these perturbations to 'reconstruct' the low elbow of the early rise and its intended alignment" ([3, pp. 222–223]).

The goals of the current study were twofold: 1. to examine the availability of intonational cues to word segmentation in the presence of segmental perturbations (F0 discontinuities), 2. to evaluate a recent proposal for the processing of F0 discontinuities ([2]). We considered the case of French accentual phrase-initial F0 rises (APRs).

1.1. French intonational structure

French accentual phrases (APs) that are not utterance-final are typically realized with a F0 rise that peaks late in the last syllable of the AP. This late rise (final rise) is accompanied by syllable lengthening and signals the primary accent. 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 (early rise or initial rise), signaling the secondary accent. Together, they form a two-rise intonation pattern LHLH, see Figure 1, top panel. In short APs, a single rise combining the characteristics of the late rise and early rise is often realized and is hence referred to as the L(HiL)H* by [7] or "the single-rise L1H2" by [1], see Figure 1, bottom panel.

1.2. Intonational cues to word segmentation and lexical access in French

Offline identification tasks were used in [3]/[4] to show that French listeners use the early F0 rise of the two-rise LHLH pattern to locate the beginning of content words. These results were extended by [8] and [1], using both offline and online tasks to show that the AP-initial boundary rises (APRs) described in §1.1 not only serve as cues to word segmentation, but also speed lexical access.

One limitation of [3]/[4] was that the materials were restricted to items with liquid and nasal consonants in the critical region. This assured a smooth F0 curve and facilitated F0 resynthesis. In subsequent studies ([8]/[1]), we used items with more segmental variety, containing not only liquids and nasals, but also voiceless obstruents in the critical region. Segmental composition, however, was neither directly manipulated, nor examined as factor in these studies. In the current study, we reanalyze the results of [1], considering as a factor whether the critical region contains a consonant that leaves the F0 curve intact or a consonant that leads to a gap in the F0 curve, as well as what kind of gap is created.

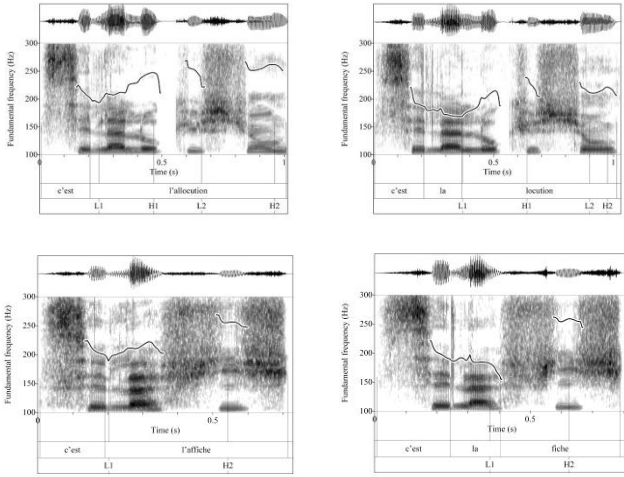


Figure 1: Example pairs of phonemically identical pairs of items from [8]/[1]. Top: l'allocation 'the speech' (left), la locution 'the expression' (right). Bottom: l'affiche 'the poster' (left), la fiche 'the sheet (of paper)' (right).

1.3. The processing of F0 discontinuity

There is a small but growing body of research on how listeners process F0 discontinuity (e.g., [9], [10]), whether they "fill in" gaps, ignore them or process them in other ways. We examine the use of intonational cues to word segmentation in French with respect to the recent proposal of [2].

These authors provide evidence from the perception of German that listeners process gaps in the F0 curve differently depending on two factors: 1. what restoring the missing F0 values would require of the listener and 2. the nature of the gap in F0. If a gap interrupts a relatively straight stretch, restoring the missing F0 requires *interpolation* between points. If a gap occurs in a region where a F0 peak or low "elbow" (a "turning point") would otherwise be, restoring the missing F0 requires *extrapolation* or the projection of the existence of a point that is not physically present. That is, in the former case, listeners can simply "connect the dots"; in the latter, they also need to add back in a missing dot. In both cases, the gap in F0 may be (relatively) silent, as in a perturbation due to a voiceless stop, or it may be "filled" by the frication noise of a fricative. A schematic representation of these different kinds of gaps is given in Figure 2.

Unlike silent gaps, gaps filled by frication noise are basically able to create pitch impressions in listeners. For example, as [11, p. 157] states: "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". Until recently, such aperiodic pitch impressions were only considered relevant in psychoacoustic contexts ("sibilant pitch", [12]) or in whispered speech ([13]). However, studies by [14], [15], [16] showed that speakers vary the spectral energy distributions of fricatives such that the pitch impressions they create in listeners reflect the adjacent F0 context. In addition, evidence from cross-splicing perception experiments suggest that listeners integrate the pitch impressions of fricative noises with the pitch contour based on F0 ([17], [18]). This relevance of fricative noises in the production and perception of pitch patterns in speech was termed "segmental intonation." Recent

studies refined this notion of "segmental intonation" ([19]) and replicated it for other languages ([20][21]).

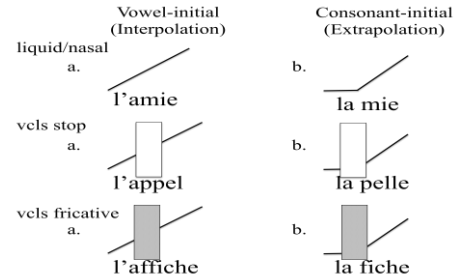


Figure 2: Schematic representation of F0 contours with three types of critical consonants 1. a liquid/nasal (relatively smooth, uninterrupted F0 curve), 2. a voiceless stop (silent gap in the F0 curve), and 3. a voiceless fricative (filled gap in the F0 curve). The (a) examples involve interpolation or "connecting the dots"; the (b) examples involve extrapolation or adding back in a missing dot.

In the model of [2], "segmental intonation" predicts that, in terms of listener judgments, conditions in which gaps are filled by voiceless frication will be more similar to sonorant-consonant (here liquid/nasal-consonant) than to voiceless-stop conditions. Additionally, the model predicts with reference to psychoacoustic evidence that listeners can interpolate but not extrapolate F0 discontinuities, regardless of what type of sound causes this discontinuity.

The results of [2] showed that "[l]isteners do not fill in all F0 gaps, but they seem to fill in some (non-silent) F0 gaps" (p. 234). A gap filled by fricative noise allowed listeners to interpolate missing F0 (across the straight stretch of a rise), but not to extrapolate the presence of a missing peak. In contrast, no evidence was found that listeners filled in silent gaps caused by voiceless stops. In [2], tonal scaling and prominence perception were used to test the model. It is, however, a general model of intonation. As [2] write: "treating unvoiced fricatives and plosives differently is not just important for modeling prominence but also for modeling intonation" (p. 234). The model therefore makes testable predictions about how different kinds of F0 discontinuities interfere with cues to word segmentation in French accentual phrases.

2. Method

2.1. Materials and Procedures

The study in [8] used 30 phonemically identical sequences such as *l'affiche/la fiche* and *l'allocation/la locution*. The first member of each pair contains 1. the elided form of the definite article (*l'*) followed by a vowel-initial noun (*affiche, allocation*) (the *vowel-initial* segmentation condition), and the second contains 2. a full form of the definite article (*la*) followed by a consonant-initial noun (*fiche, locution*) (the *consonant-initial* segmentation condition). Since the early F0 rise starts at the beginning of a content word, it occurs earlier in *l'affiche* and *l'allocation* (at the first syllable ([la]) than in *la fiche* and *la locution* (at the second syllable ([fɛʃ], [lo])). This difference in tonal alignment is illustrated in Figure 1.

The [8] study used natural read speech stimuli and a variety of offline and online tasks to examine the influence of intonational cues to word segmentation and lexical access.

Although the results supported the use of such cues, phonetic analyses showed differences not only in F0, but also in formant structure and duration between the vowel-initial (*l'affiche*) and consonant-initial (*la fiche*) items. For example, when the first syllable [la] was content-word initial (in the vowel-initial condition), this syllable was longer and the vowel ([a]) had a lower F2. To control for the influence of these confounding factors and to isolate that of F0, [1] resynthesized the F0 of the natural speech stimuli from [8].

For each experimental pair in [1], there were five versions, corresponding to five experimental conditions: "(1) natural-consonant (e.g., *la fiche*, consonant initial production, no resynthesis), (2) natural-vowel (e.g., *l'affiche*, vowel-initial production, no resynthesis), (3) resynth-consonant-equal (e.g., consonant-initial item *la fiche* with f0 of the first vowel /a/ resynthesized to equal that of the vowel-initial item *l'affiche*), (4) resynth-consonant-higher (e.g., consonant-initial item *la fiche* with the f0 of the first vowel /a/ resynthesized to be higher than that of the vowel-initial item *l'affiche*), and (5) resynth-consonant-lower (e.g., consonant-initial item *la fiche* with the F0 of the first vowel /a/ resynthesized to be lower than that of the consonant-initial item *la fiche*)" (p. 778). As expected, [1] found that increasing the F0 in the /a/ of the consonant-initial production (*la fiche*) gave rise to more vowel-initial segmentation (*l'affiche*), in the resynth-consonant-equal condition and even more in the resynth-consonant-higher condition. (However, decreasing the F0 value in the /a/ of the consonant-initial production (*la fiche*) did not give rise to more consonant-initial segmentations.)

In the current study, we assigned each pair of items from the materials of [1] to one of three conditions, based on the onset of the second syllable and whether this critical consonant (or consonant cluster) led to a discontinuity in the F0 curve in the critical region: 1. *Liquids/nasals* with continuous F0: lateral liquids (e.g., *l'allocation/la locution*) and nasals (e.g., *l'amie/la mie*), 2. *Voiceless fricatives* with a gap in the F0 curve "filled" by frication noise (e.g., *l'affiche/la fiche*). 3. *Voiceless stops* with a silent gap in F0 curve, including simple stops (e.g., *l'appel/la pelle*) and stop-liquid clusters (e.g., *l'attraction/la traction*). Note that our criterion was phonetic rather than phonological. For example, in French, the rhotic liquid patterns phonologically with the lateral liquid as a sonorant, but in the materials was always produced as a fricative (/ʁ/, e.g. in *l'arène/la reine*) in the critical region. We excluded from the analysis one pair (*l'annotation/la notation*) for which the vowel-initial segmentation item was produced with an unexpected F0 pattern. We also excluded the five pairs of items with phonologically voiced fricatives in the critical region (e.g., *l'avenue/la venue*). The processing of voiced fricatives is an interesting question, but in these materials, one or both members of pairs with phonologically voiced critical consonants were produced with partial or full devoicing. These items therefore did not form a phonetically homogenous category. There were no voiced stops in the materials. In total, 22 pairs remained for the analyses, unevenly distributed across conditions: six pairs of liquids/nasals, four pairs of voiceless fricatives, and 12 pairs of voiceless stops.

We reanalyzed the results of [1] with respect to the two segmentation (consonant-initial, vowel-initial) and the three consonant (liquids/nasals, voiceless fricatives, voiceless stops) conditions to see whether the overall pattern of results

held across conditions or differed in the directions predicted by [2].

Experiment 1 of [1] was an offline forced choice identification task (a replication Experiment 2 of [8]). Participants heard one member of a pair embedded in a neutral carrier and identified the noun by making a forced choice between two possibilities (e.g., *affiche* vs. *fiche*).

2.2. Predictions

For the [1] materials, the model proposed in [2] makes a number of testable predictions:

Prediction 1. Liquid/nasal critical consonants (*l'amie/la mie*) support the intonational cue to a content word beginning in both the vowel-initial (interpolation) and the consonant-initial (extrapolation) segmentation conditions. The liquid/nasal consonant items have an intact F0 curve with no gap in the critical region. (This and the following descriptions are schematized in Figure 2.) The entire intonational cue to word segmentation (including both the beginning and the end of the F0 rise) is thus preserved and is available to help listeners identify the intended segmentation.

Prediction 2. a. Fricative critical consonants will support this intonational cue to segmentation, but only in the vowel-initial (interpolation) segmentation condition (*l'affiche*), where the fricative noise of the critical consonants allows the listener to bridge the gap and "fill in" the missing F0 values. **b.** In the consonant-initial (extrapolation) segmentation condition, however, voiceless fricative critical consonants (*la fiche*) will not support the intonational cue. Voiceless fricatives create a gap (albeit a "filled" gap) and crucially a missing turning point that the listener cannot restore. This will degrade the quality of the intonational cue: the beginning of the rise is missing, although the end (the peak) is preserved.

Prediction 3. a. In the vowel-initial (interpolation) segmentation condition (*l'appel*), the presence of a voiceless stop critical consonants will also degrade the intonational cue, but for a different reason. In [2] clearer prominence differences were observed when more of the F0 peak was present. We expect a similar difference here. In this condition, while both the beginning (elbow) and the end (peak) of the rise are preserved, the voiceless stop creates a silent gap in the middle of the F0 rise and therefore less of the F0 rise is physically present. **b.** In the consonant-initial (extrapolation) segmentation condition, voiceless stop critical consonants (*la pelle*) will not support the intonational cue. The quality of the intonational cue will be degraded for the same reasons as for voiceless fricatives in condition 2b.

Support for the intonational cue to word segmentation should be reflected in good identification performance.

3. Results

Given the uneven distribution of critical consonant types in the [1] items, we give a description of the observed patterns, rather than attempting a statistical analysis with little power. With very few pairs in the liquid/nasal (six) and fricative (four) conditions (vs. 16 in the stop condition), the patterns may reflect inter-subject variability rather than a true difference.

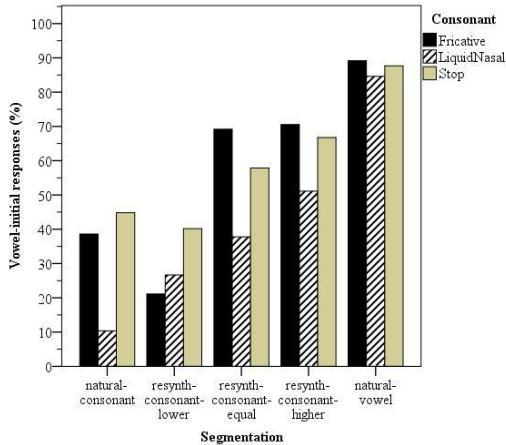


Figure 3: Spinelli et al. (2010) Segmentation task: Vowel-initial responses by segmentation and critical consonant type

The results of the offline identification task in [1] (Expt. 1) are given in Figure 3. The dependent variable is percentage of vowel-initial responses (e.g., *affiche*).

Considering first the two natural productions (natural-consonant-initial and natural-vowel-initial), for each of the three consonant conditions (liquid/nasal, fricative, stop), the natural-vowel-initial (interpolation) stimuli received a comparable high percentage of vowel-initial responses (85–89%). This was expected for the liquid/nasal and fricative critical consonants (Predictions 1 and 2a). According to Prediction 3, in the stop critical consonant condition, the silent gap should degrade the effectiveness of the cue of the early rise. As expected, the natural-consonant-initial (extrapolation) stimuli received a much lower percentage of vowel-initial responses, but the percentage varied according to consonant condition, from 10.4% for liquid/nasal critical consonants to 44.8% for stops and 38.6% for fricatives. Cues to word segmentation are strong for items with liquid/nasal critical consonants, in line with Prediction 1. The poorer performance for fricative and stop critical consonants (for stops, close to chance) is in line with Predictions 2b and 3b. It is, however, surprising since in the natural speech stimuli, all cues are present (formant structure, duration, F0, etc.). The poor performance despite this may be due to a greater weighting for F0 cues that for other cues. A re-analysis by consonant type of the identical offline identification task for the same experimental pairs but different listeners in [8] (Experiment 2) showed similar patterns, in particular for the consonant-initial condition, for which liquids/nasals give better cues. Note that the study in [8] used only natural speech stimuli.

To isolate the influence of intonational cues, we first compared the natural-consonant-initial condition and the resynth-consonant-equal condition. Crucially, there were more vowel-initial choices in the resynth-consonant-equal condition (as in [1] for all experimental items). Raising the F0 of the /a/ in natural consonant-initial productions to a value compatible with that of a vowel-initial segmentation led to more vowel-initial responses. This difference is more marked for items with liquid/nasal critical consonants (10.4% to 37.8%) than for those with stop critical consonants (44.8% to 57.9%), in line with Predictions 1 and 3. There was a further gain when the F0 of the /a/ was raised even higher (resynth-consonant-higher), a difference again more marked for items with

liquid/nasal and fricative critical consonants (10.4% to 51.1% and 38.6% to 70.6%, respectively) than for those with stop critical consonants (44.8% to 66.8%), in line with Predictions 1 and 3a. The great increase in vowel-initial responses for fricative critical consonants is also in line with Prediction 2a, although given the very small number of fricative pairs, we must interpret this result with caution.

4. Discussion and Conclusions

Results of this exploratory analysis provide some evidence that segmental perturbations degrade intonational cues to content word beginnings in French. In an identification task using resynthesized stimuli, an F0 rise early in the AP cued content word beginnings for items with all three types of critical consonants (liquids/nasals, fricatives, and stops). Liquids/nasals, however, for which the F0 curve is smooth and uninterrupted, seem to provide stronger cues than fricatives and stops. But even for stops, part of the cue is preserved (the peak and a portion of the rise), even though the F0 turning point corresponding to the beginning of the rise is not present.

The imbalance in the numbers of items in the three critical consonant conditions, with many fewer items for liquids/nasals and fricatives than for stops, prevents us from making firm conclusions. We have, however, found empirical evidence in this preliminary investigation that it is worth taking a closer and more systematic look at the questions raised. In a new study, we will explicitly manipulate critical consonant as a factor. This will also allow us to further substantiate Niebuhr's notion of "segmental intonation" ([14], [16]), according to which spectral differences in the realization of a fricative help listeners to bridge the gap and "continue" the intonation pattern, even in the absence of F0.

The precise nature of the intonational cues to word segmentation in French remains an open question. One possibility is that "the early rise is a two-part cue to a content word beginning: a L tone followed by a H tone, a treatment in line with an AM [autosegmental-metrical] model. The cue is most effective when both parts are present (i.e., when there is not only a L1, but also a following H1), but listeners can make use of the cue even when only one is present..." ([4, p. 45]). Another possibility is that the cue is stronger when more F0 is physically present ([2]). A third possibility is that the observed differences can be accounted for something like the Tonal Center of Gravity (TCoG) "a *gestalt* or global measure of F0 event localization" that "focus[es] not on the onsets and offsets of pitch movements but rather on the overall distribution of the 'mass' or bulk of raised F0 in both time and frequency space" and explicitly seeks to characterize intonation patterns where turning points are missing due to segmental perturbation and other factors ([22][10]). However, in an examination of the [8]/[1] materials with respect to the TCoG, [23] were unable to distinguish vowel-initial and consonant-initial members of the experimental pairs using this measure. Our more recent attempts varying the TCoG region of interest have also been unsuccessful in the distinguishing the members of the pairs.

5. Acknowledgments

We thank Elsa Spinelli for her assistance. This research is funded by a French National Research Agency award to Corine Astésano (ANR-12-BSH2-0001-01, "PhonIACog", 01/01/13–12/31/16).

6. References

- [1] E. Spinelli, N. Grimault, F. Meunier, and P. Welby. "An intonational cue to word segmentation in phonemically identical sequences," *Attention, Perception and Psychophysics* vol. 72, pp. 775–787, 2010.
- [2] H. Mixdorff and O. Niebuhr. "The influence of F0 contour continuity on prominence perception," in *Interspeech 2009 – 14th Annual Conference of the International Speech Communication Association, Lyon, France*, pp. 230–234, 2013.
- [3] P. Welby. *The Slaying of Lady Mondegreen, being a Study of French Tonal Association and Alignment and their Role in Speech Segmentation*. Ph.D. dissertation, The Ohio State University, 2003.
- [4] P. Welby. "The role of early fundamental frequency rises and elbows in French word segmentation," *Speech Communication*, vol. 49, pp. 28–48, 2007.
- [5] S. Kim and T. Cho. "The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean," *Journal of the Acoustical Society of America*, vol. 125, pp. 3373–3386, 2009.
- [6] N. Warner, T. Otake, T., and T. Arai, "Intonational structure as a word-boundary cue in Tokyo Japanese," *Language & Speech*, 53, pp. 107–131, 2010.
- [7] S.-A. Jun and C. Fougeron, "Realizations of accentual phrase in French," *Probus*, vol. 14, pp. 147–172, 2002.
- [8] Spinelli, E., P. Welby, and A.-L. Schaegis, "Fine-grained access to targets and competitors in phonemically identical spoken sequences: The case of French elision," *Language and Cognitive Processes*, vol. 22, pp. 828–859, 2007.
- [9] A. S. Bregman. *Auditory Scene Analysis*. Cambridge, MA: M.I.T. Press, 1990.
- [10] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel. "Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology," *Laboratory Phonology* vol. 3, pp. 337–383, 2012.
- [11] K. Johnson. *Acoustic and Auditory Phonetics*, Oxford: Wiley-Blackwell, 2012.
- [12] H. Traunmüller, "Some aspects of the sound of speech sounds," in Schouten, *The Psychophysics of Speech Perception*, pp. 293–305, Dordrecht: Nijhoff, 1987.
- [13] W. Meyer-Eppler, "Realization of prosodic features in whispered speech," *Journal of the Acoustical Society of America*, vol. 29, pp. 104–106, 1957.
- [14] O. Niebuhr. "Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German," *Journal of the Acoustical Society of America*, vol. 124, pp. 1252–1263, 2008.
- [15] O. Niebuhr. "Intonation segments and segmental intonations," in *Interspeech 2009 – 10th Annual Conference of the International Speech Communication Association*, Brighton, England, pp. 2435–2438, 2009.
- [16] O. Niebuhr. "At the edge of intonation – The interplay of utterance-final F0 movements and voiceless fricative sounds," *Phonetica* vol. 69, pp. 7–27, 2012.
- [17] O. Niebuhr, "Drawing on intonation drawings - An alternative approach to the perception of pitch accents and phrase-final intonation movements," *KALIPHO* vol. 1, pp. 1–40, 2011.
- [18] W. F. L. Heeren, "Coding pitch differences in voiceless fricatives: Whispered relative to normal speech," *Journal of the Acoustical Society of America* vol. 138, pp. 3427–3428, 2015.
- [19] S. Ritter and T.B. Roettger. "Speakers modulate noise-induced pitch according to intonational context," in *7th International Conference on Speech Prosody*, Dublin, Ireland, pp. 890–893, 2014.
- [20] Żygis, M., D. Pape, L. Jesus, and M. Jaskula, "How do voiceless fricatives contribute to intended intonation? A comparison of whispered, semi-whispered and normal speech," in *10th International Seminar on Speech Production*, Cologne, Germany, pp. 472–475, 2014.
- [21] M. Żygis, D. Pape, L. Jesus, and M. Jaskula, "The effects of intonation on acoustic properties of fricatives," in *18th International Congress of Phonetic Sciences*, Glasgow, Scotland, pp. 1–4, 2015.
- [22] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel. "The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation," in *Speech Prosody 2010*, pp. 1–4, 2010.
- [23] P. Welby, R. Espesser, and C. Meunier. "In search of intonational cues to content word beginnings in conversational speech," in *New Tools and Methods for Very-Large-Scale Phonetics Research*, Philadelphia, United States, 2011.