Speakers modulate noise-induced pitch according to intonational context

Simon Ritter, Timo B. Roettger

I*f*L Phonetik, University of Cologne

{simon.ritter; timo.roettger}@uni-koeln.de

Abstract

Recent studies have shown that speakers systematically modulate properties of voiceless segments according to intonational context. More specifically, in the absence of fundamental frequency (F0), speakers appear to adjust the Center of Gravity (CoG) and the intensity of voiceless fricatives to convey the impression of pitch. In line with these findings, the present production study extends earlier work and investigates noise-induced properties of fricatives, modulated by the intonational context. It is shown for German that the mean CoG and intensity of intended contours with a high boundary tone are higher than those produced for intended contours with a low boundary tone. Furthermore, looking at the development of CoG and intensity over the time course of the fricative, the trajectories corresponding to the boundary tones differ in intercept (CoG and intensity) and slope (intensity), i.e. reveal a steeper fall in case of a corresponding falling tone.

Index Terms: noise-induced pitch, intonation, boundary tone, truncation

1. Introduction

Segmental and suprasegmental properties of the speech signal have traditionally been described as two separate levels. More specifically, intonation has been mainly associated with the acoustic parameter of fundamental frequency (F0), which is superimposed on the segmental string to bear different communicative functions. In intonation research, segments themselves are traditionally considered to constitute a potential perturbation of F0. In particular, voiceless parts of the signal have been regarded as communicatively irrelevant for the interpretation of the meaning conveyed by the F0 contour. However, intonation contours are frequently interrupted by the lack of voiced segments. As a consequence, speakers realize F0 movements on the segmental material available, i.e. they might truncate the contour (e.g. [1]). In the case of truncation, the contour simply ends earlier. Work on German phrase-final intonation patterns provides evidence for truncation in nuclear contours consisting of a high peak followed by a fall [2]: If the voiced material available at the end of the phrase was limited (e.g. in words like "Schiff" /ʃɪf/), the falling intonation movement is not realized completely (truncated). In some cases, the fall is entirely absent. Despite missing F0 information, speakers appear to have no problem with understanding each other's communicative intentions. In fact, [2] notes that even if the final fall is missing in the F0, German listeners perceive the "word as having 'falling pitch'" [2:140].

So why can speakers understand each other even though communicative relevant F0 movements may be entirely absent? One possible answer to this question is that the signal is over specified: acoustically different parts of the signal can serve as cues for certain meanings. In particular, other parts of the signal might function as acoustic cues which correspond to the intended meanings.

Experiments with whispered speech have shown that segmental cues can be used to convey meanings that would otherwise be encoded in the F0. Listeners of Mandarin compensate for the lack of F0 in whispered speech e.g. by using duration of the syllable as an auditory cue for the contrast between lexical tones [3]. Such a usage of acoustic substitutes for F0 is not restricted to whispered speech: Recent experiments have shown that speakers consistently adjust noise-induced differences to intonational contexts in normal speech. Several experiments demonstrated that frication and aspiration noise of phrase medial and final voiceless obstruents corresponds to the expected modulation of F0 [4,5,6,7]. In particular, voiceless parts of the signal corresponding to high/rising tones exhibit higher mean Center of Gravity (CoG) and higher mean intensity values than their counterparts corresponding to low/falling tones.

Niebuhr [7] investigated polar questions and statements in German. German polar questions typically end in a rise, whereas statements typically end in a fall (in Autosegmental-Metrical terms H-% and L-%, respectively [8]). In his study, target words ending in voiceless fricatives were placed at the end of the utterance. Acoustic measurements revealed that fricatives obtained higher CoG and intensity means at the end of questions than at the end of statements. Niebuhr interpreted these differences as noise-induced correlates of the intended boundary tone. This conclusion, however, is limited to some degree: The contours not only differ in their boundary tones, they are characterized by the combination of a boundary tone and a particular pitch accent type preceding the boundary tone. In the case of the question it is a low pitch accent (L*), in the case of the statement it is either a high (H*), a rising (L+H*) or a falling (H+L*, H+!H*) pitch accent. Thus, the factor boundary tone is confounded with the pitch accent type, making a conclusion towards a direct causal link between missing F0 information and spectral differences difficult.

This is important because a considerable amount of the global contour differences between statements and questions is manifested through the pitch accent preceding the boundary tone, i.e. the F0 movement of the pitch accent is realized within the vowel. This leaves the listeners with relevant information of the tonal movement encoded directly in the F0 to distinguish sentence modalities. In turn, modulation of spectral characteristics of the fricatives might be less required to signal the contrast. To shed further light on the relation between boundary tone and noise-induced characteristics of voiceless sounds, we report on a production study on German boundary tones extending Niebuhr's findings. In this study, we keep the pitch accent type constant and vary only the boundary tone.

A further contribution of this paper is the nature of our dependent variable. Tonal events are inherently dynamic, i.e. a function of F0 developing over time. Previous work only looked at summary measures, which reflect the averages over the whole segment, i.e. mean CoG and intensity. However, CoG and intensity may change dynamically throughout the duration of a segment. Any differences of CoG or intensity found for the arithmetic mean may reflect (a) a baseline

10.21437/SpeechProsody.2014-166

difference of CoG (an overall difference with a similar trajectory), (b) a trajectory difference starting at the same intercept or (c) a combination of an intercept difference and a trajectory difference. A first indication of dynamic differences of such noise-induced cues was reported by [4]. CoG and intensity were measured at the beginning and the end of the aspiration of /t/. The results showed that intonational contexts with a final F0 fall exhibited steeper slopes of the highest spectral energy in the aspiration. In the present study, we explore the development of CoG and intensity over the time course of the fricative in addition to the static mean.

2. Methodology

2.1 Reading material

Six monosyllabic nouns with CVC structures served as target words with three words for each of the target fricatives (listed in table 1). The target sounds were the postalveolar and uvular voiceless fricatives (/ʃ/ and /χ/). Following [7]'s findings, those sounds were chosen in order to elicit the strongest segmental pitch effects.

/ʃ/	Fisch ('fish')	/fɪʃ/
	Tisch ('table')	/tɪʃ/
	Busch ('bush')	/bʊʃ/
/χ/	Koch ('cook')	/kɔχ/
	Loch ('hole')	/lɔχ/
	Bach ('stream')	/baχ/

Table 1: *Target words sorted by fricative*

The vowels in syllable nucleus position were phonologically short. Short vowels were chosen in order to reduce the voiced material available to realize the F0 movements before the voiceless fricatives at the end of the target words.

In addition to the six target words, we used two control words. Both contained voiced segments only. One was monosyllabic (*See* 'lake' /ze:/) and one was trisyllabic (*Brombeere* 'blackberry' /'bʀɔmbe:ʀə/). We included these control words to elicit undisturbed F0 contours for the contexts under scrutiny.

The target words were embedded in short dialogues on everyday topics. The dialogues comprised two to three turns per speaker. Each target word occurred in two different contexts: In context A, the target word was the first noun in an *enumeration* of three or more nouns. In context B, the target word was utterance final in a *contrastive focus* statement (corrective contrast: "is it X?" "No it is *Y.*").

The syntactic structures of the critical utterances as well as the semantic-pragmatic context frames set by the preceding utterances were designed in such a way that they elicited fundamentally different types of edge contours: Context A mainly elicited a high nuclear pitch accent (L+H* or H*) on the target followed by a high boundary (H- or H-%) resulting in a plateau. Context B mainly elicited a high nuclear pitch accent (L+H* or H*) on the target with a terminal falling utterance-final movement (L-%). This contour is typical for a contrastive focus statement in German.

2.2 Participants and recording procedure

Twelve Participants (mean age = 22; 6 men; 6 women) were seated in front of a computer screen together with one of the experimenters and read aloud the mini-dialogues at a time. They were instructed to read each dialog silently first. After the silent reading they read the dialog together with the experimenter. They were instructed to read the contexts as naturally as possible.

2.3 Analyses

The recordings were digitized at a sampling rate of 44.1 kHz (16bit). All acoustic material was manually annotated. For the acoustic analysis of the segments, we identified segmental boundaries of the target word using a waveform and a wide-band spectrogram. All segmental boundaries of vowels and consonants were labeled at abrupt changes in the spectra. Intonation contours were labeled according to the GToBI annotation system [8]. Those productions of the target utterances that could not be counted as instances of one of the two contours described in §2.1 were excluded from the analysis (n = 29).

Based on the labels for the acoustic boundaries, we measured the fricative *Center of Gravity* (CoG) and *intensity*. The CoG measurements were taken on the basis of spectral slices in Praat [9]. The slices resulted from a 20 ms Hamming window and were shifted in 5 ms steps across the fricative. Of each interval, CoG measurements were taken within a frequency range of 0.5-10 kHz. The frequency range covers the main spectral characteristics of /ʃ/ and /χ/ and excludes potential F0 residuals as well as high-frequency ambient noise. Using a fixed window width resulted in a different number of data points for different segment durations (e.g. longer fricatives yield more windows). To get an equal number of data points for all tokens, we normalized the data by calculating nine normalized time points of the CoG trajectory over the fricative for each token separately. In line with [5,7,10], we assume that CoG is a suitable estimate of perceived fricative pitch. For intensity, we extracted ten normalized time points of the intensity trajectory over the fricative for each token separately (due to strong perturbations of the intensity measurements at the vowel-consonant transition, the first time point was excluded from the subsequent analyses, resulting in nine time points, analogously to the CoG measurements). Measuring the CoG and intensity at different points in time throughout the segment enabled us to analyze the time course of the spectral and intensity changes.

All data were analyzed with generalized linear mixed models using *R* [11] and the package *lme4* [12]. For CoG and intensity mean (mean of the intervals), we used a Gaussian error distribution (assuming normality). We adhered to the random effect specification principles outlined in [13] including a term for random intercepts for speakers and words, which quantifies by-speaker and by-words variability. The critical fixed effects in question were BOUNDARY TONE (i.e. H vs. L) and FRICATIVE (i.e. /ʃ/ vs. /x/), and for these fixed effects, we included random slopes for speakers and words (this quantifies by-speaker and by-word variability in the effects of BOUNDARY TONE and FRICATIVE). We tested whether the inclusion of the fixed effects BOUNDARY TONE and FRICATIVE did improve the model's prediction significantly for CoG and intensity mean via likelihood ratio tests (LRT).

To test whether the actual trajectories of CoG and intensity development throughout the fricative differed as a function of time, we performed a *Growth Curve Analysis* [GCA, 14]. GCA is a multilevel regression technique designed for analysis of time series data. It fits trajectories to multilevel polynomial curves and allows for comparison of such curves. We decided

to model CoG and intensity trajectories as second order polynomials (parabola shaped curves). Thus, for both CoG and intensity, the nine time steps entered the analysis as a second order orthogonal polynomial fixed effect (including first order polynomial). The crucial effect of interest was the interaction of BOUNDARY TONE with the FIRST and SECOND ORDER POLYNOMIAL. We included a term for random intercepts for speakers and words as well as random slopes for speakers and words for the FIRST and SECOND ORDER POLYNOMIAL interaction with BOUNDARY TONE. For all models, p-values were generated using likelihood ratio tests.

3. Results and Discussion

Figure 1 shows representative contours for both conditions, as produced by one male speaker. In the upper panel, the contours over the utterance "Wir hatten Brombeere" ('We had blackberry') are displayed. Here the difference between the conditions is clearly visible on the target word "Brombeere": A high accent on the initial syllable (marked in grey) is followed by a *low tone* in the case of the contrastive statement (left), or by a *high tone* in the case of the enumeration (right). In the lower panel, where the same contours are produced over the sentence "Wir brauchen 'nen Tisch" ('We need a table'), it can be seen that the contours on the target word "Tisch" are severely truncated. This observation is in line with [2], i.e. the intended boundary tone (H-% vs. L-%) is at least highly impoverished. In this context, in which F0 is drastically reduced, we expect spectral cues to retain the contrast between the two intonation contours and their meanings.

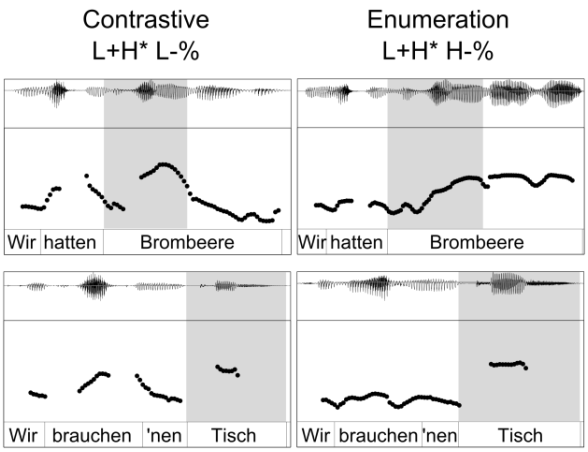

Figure 1: *F0 contours for the utterances containing the control word "Brombeere" (top) and the target word "Tisch" (bottom) in the two conditions Contrastive Focus (left) and Enumeration (right). The accented syllable is in grey.*

Mean CoG and intensity values are displayed in Figure 2. For both intensity and CoG mean, there was a significant effect of fricative, such that uvular fricatives /x/ had a 7.4 dB lower intensity (β=7.2 dB, SE=1.7, $\chi^2(1)=11.5$, $p<0.0007$) and a 1612 Hz lower CoG mean (β=1629.3 Hz, SE=241.3, $\chi^2(1)=16.43$, $p<0.0001$) than postalveolar fricatives /ʃ/. Crucially, there was a significant effect of BOUNDARY TONE on mean CoG and intensity, such that H tones elicited 4.0 dB higher mean intensities (β=4 dB, SE=1.2, $\chi^2(1)= 8.2$, $p=0.00414$) and 366.3 Hz higher CoG means (β=315.6 Hz, SE=120.2, $\chi^2(1)= 5.66$, $p=0.0174$) than L tones.

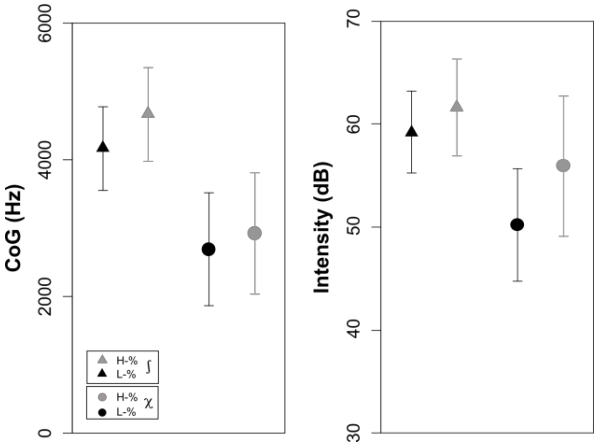

Figure 2: *CoG and intensity means and standard deviations for both fricatives and boundary tones (black = low boundary; grey = high boundary).*

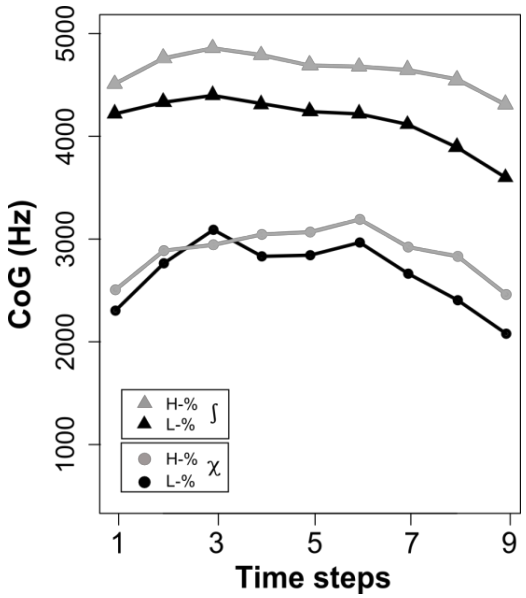

Figure 3: *CoG development as a function of time for both fricatives and boundary tones. (black = low boundary; grey = high boundary).*

To investigate the development of the measurements over time we performed growth curve analyses. As can be seen in Figure 3 and 4, for both CoG and intensity there are intercept differences, i.e. CoG and intensity start lower in the case of L-%. This difference appears to become stronger over time for intensity, that is, intensity trajectories have slightly steeper slopes for L-%. Looking at the CoG trajectories, there appear to be only small slope differences between L-% and H-%, mainly manifested in the last three time steps.

This is reflected in a significant interaction effect of BOUNDARY TONE with the FIRST ORDER POLYNOMIAL (the linear component of the models) for intensity ($\chi^2(1)=14.2$, $p=0.0002$) such that L tones corresponded to intensity trajectories with a steeper negative slope. This interaction is not significant for CoG ($\chi^2(1)=2.2$, $p=0.14$), although numerical trends point towards a comparable slope difference. Even though the SECOND ORDER POLYNOMIAL (the square components of the models) ($\chi^2(1)<0.35$, $p>0.55$) did not significantly interact with BOUNDARY TONE there were numeric tendencies suggesting that the trajectories elicited by H-% are slightly less curved than the trajectories elicited by L-%, or in other words: flatter. It is important to note that even though we found significant differences between the conditions, CoG and intensity decreases over time for both low and high boundary tones.

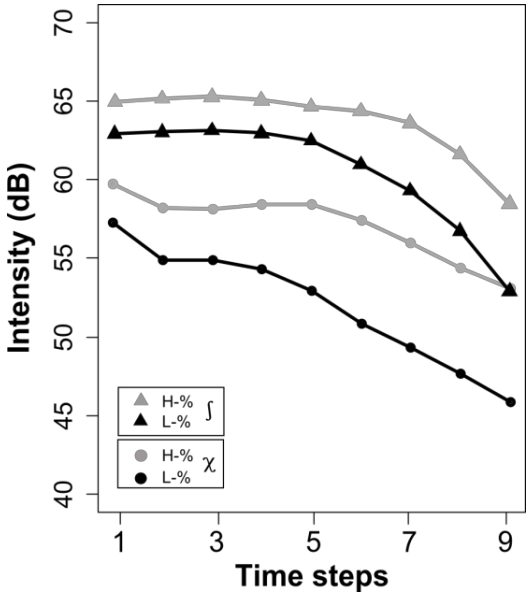

Figure 4: *Intensity development as a function of time for both fricatives and boundary tones. (black = low boundary; grey = high boundary).*

The present study has demonstrated that German speakers systematically modulate properties of voiceless segments to convey the meaning encoded in the intonation contours. Specifically, we have found that the spectral properties of voiceless fricatives, as reflected in the measure of CoG and intensity, are modulated in different intonational contexts: Higher CoG and intensity means are found for fricatives at the end of phrases ending with a high boundary tone. Thus, this study replicates earlier findings reported by [7], circumventing the confounding of boundary tone and pitch accents, and confirms that German speakers produce noise-induced correlates of intended boundary tones [4,5,6,7].

Crucially, the obtained mean differences can be ascribed to differences in development over time, at least for intensity: Intensity of high boundary tones starts higher (higher intercept) and remains flatter (flatter slope, less curved) than those of low boundary tones. CoG differences appear to be mainly due to intercept differences, i.e. CoG for high boundary tones starts higher than those of low boundary tones with a comparable development of the trajectory over time.

4. General Discussion

Grabe [2] noted that German native listeners hear truncated contours as falling even if an explicit fall in the F0 is missing. So listeners infer communicative intention even though relevant F0 movements may be entirely absent. This might be possible due to a highly over specified signal. The present study demonstrates that voiceless parts of the signal, which are not able to convey F0 information, bear their own acoustic dimensions which correspond to the missing F0 information. The question arises as to whether these acoustic differences have any communicative function – in other words, whether speakers are able to use these cues to distinguish e.g. sentence modalities. A semantic differential task performed by [4] has demonstrated that noise-induced cues of /t/ aspiration were able to shift the attitudinal meaning of the stimuli towards the meaning profile of the respective intonational context. This is a first indication that speakers, indeed, might be able to use such subtle cues communicatively. Future research is needed to further elaborate the impact of noise-induced pitch on communication.

Generally, research into this phenomenon would benefit from cross-linguistic investigations. It is important to note that the acoustic differences reported here are very subtle and prone to variation necessarily limiting its examination. Other languages could in fact be better suited for such investigations. For example, Tashlhiyt Berber, an Afroasiatic language spoken in Morocco, can have whole utterances with neither a vowel nor a voiced segment. As a result, the phonetic opportunity it affords for the execution of intonational pitch movements is exceptionally limited. In fact, it has been reported that entire complex tonal movements (Rise-Fall) can be missing in certain phonotactic environments [15,16]. Speakers of such languages might rely heavily on other cues than F0 to capture the meaning encoded in the intonation contour.

To conclude, the present findings suggest that the traditionally separated levels of analysis – segmental and suprasegmental – are strongly intertwined. Voiceless segments have been regarded as irrelevant for the interpretation of the F0 contour. They have been treated as elements to be ignored. The present findings, however, demonstrate that these parts of the signal contain acoustic dimensions that potentially contribute to the perception of the intonation contour. Thus, the results may provide an answer to the question why German listeners perceive a truncated contour as falling although there is no fall in the F0 contour [2].

5. Acknowledgements

We would like to thank Oliver Niebuhr for his valuable comments and suggestions and Bodo Winter for his statistical advice on our analyses.

6. References

[1] Erikson, Y. and M. Alstermark, "Fundamental Frequency correlates of the grave word accent in Swedish: the effect of vowel duration", Speech Transmission Laboratory, Quarterly Progress and Status Report, 2-3, KTH, Sweden, 1972.

[2] Grabe, E., "Pitch accent realisation in English and German", Journal of Phonetics, 26: 129-144, 1998.

[3] Liu, S. and A.G. Samuel, "Perception of Mandarin Lexical Tones when F0 Information is Neutralized", Language and Speech, 47: 109-138, 2004.

[4] Niebuhr, O., "Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German", Journal of the Acoustical Society of America, 124: 1252–1263, 2008.
[5] Niebuhr, O., "Intonation segments and segmental intonations", Proc. of the 10th Interspeech conference, Brighton, UK, 2435-2438, 2009.
[6] Niebuhr, O., C. Lill., J. Neuschulz, "At the segment-prosody divide: The interplay of intonation, sibilant pitch and sibilant assimilation", Proc. of the 17th ICPhS, Hong Kong, China, 1478-1481, 2011.
[7] Niebuhr, O., "At the edge of intonation: the interplay of utterance-final F0 movements and voiceless fricative sounds", Phonetica, 69: 7–27, 2012.
[8] Grice, M. and S. Baumann, "Deutsche Intonation und GToBI", Linguistische Berichte, 191: 267–298, 2002.
[9] Boersma, P, "Praat, a system for doing phonetics by computer", Glot International, 5, 341-345. 2002.
[10] Traunmüller, H. "Some aspects of the sound of speech sounds", in M. E. Schouten [Ed.], The psychophysics of speech perception, 293–305, Nijhoff: Dordrecht, 1987.
[11] R Core Team, "R: A Language and Environment for Statistical Computing", http://www.R-project.org/, 2012.
[12] Bates, D., M. Maechler, B. Bolker, "lme4: Linear mixed- effects modelsm using S4 classes", R package version: 0.999999-0, 2012.
[13] Barr, D. J., R. Levy, C. Scheepers, H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal", Journal of Memory and Language, 68: 255-278, 2013.
[14] Mirman, D., J.A. Dixon, J.S. Magnuson, "Statistical and computational models of the visual world paradigm: Growth curves and individual differences", Journal of Memory and Language, 59, 475-494, 2008.
[15] Röttger, T. B., R. Ridouane, M. Grice, "Sonority and syllable weight determine tonal association in Tashlhiyt Berber", Proc. of 6th International Conference on Speech Prosody, Shanghai, 2012.
[16] Röttger, T. B., R. Ridouane, M. Grice, "Phonetic alignment and phonological association in Tashlhiyt Berber", *Journal of Acoustical Society of America*, 133: 3572, 2013.