

# AT THE SEGMENT-PROSODY DIVIDE: THE INTERPLAY OF INTONATION, SIBILANT PITCH AND SIBILANT ASSIMILATION

Oliver Niebuhr, Cassandra Lill & Jessica Neuschulz

Department of General & Comparative Linguistics, University of Kiel, Germany  
niebuhr@linguistik.uni-kiel.de; cassandra\_lill@web.de; jessi.neuschulz@gmx.de

## ABSTRACT

The presented production study continues and expands a line of research on “segmental intonation” in German. It shows that the noise-induced pitch impressions of sibilants, measured in terms of spectral centre-of-gravity, are adjusted to the into-national context not only utterance-finally, but also utterance-medially in /sz/ and /sʃ/ sequences. In the case of /sʃ/, the degree of /s/-to-[ʃ] assimilation is exploited as a means of sibilant-pitch adjustment.

**Keywords:** sibilant pitch, assimilation, intonation

## 1. INTRODUCTION

Traditionally, the speech code is dichotomized in two coding layers: segments and prosodies. However, recent studies weakened the traditional dichotomy. For example, in everyday communication function words can entirely lose their separate *segmental* representation in favour of articulatory and phonatory *prosodies* that are superimposed on the segmental environment, and that maintain the essential sound characteristics of the function word [2, 8]. Conversely, intonation is traditionally associated with the *prosodic* parameter F0. However, what happens when F0 is interrupted by voiceless segments? It was shown by [5, 6] that, in utterance-final position, the spectral energy distributions in fricative and aspiration-noise *segments* are adapted to match in terms of their resulting pitch impressions with the high or low endpoints of rising or falling intonations. These noise pitches help identifying the communicative functions coded by the utterance-final intonation rises and falls and hence undermine the truncation concept of [1].

Our production study continues this line of research at the heart of the intersection of segments and prosodies. The study was done for German and started from the alveolar and postalveolar fricative segments [s,z] and [ʃ]. Depending on where their noise energy is concentrated in the frequency spectrum, each of them can induce ‘sibilant pitch’

within a substantial semitone range [9]; and German speakers in fact adjust these sibilant pitches to the high or low utterance-final intonation targets [6]. However, is this sibilant-pitch adjustment for some reasons a peculiarity of utterance-final sibilants, or does it also occur utterance-medially? This is our first and primary question.

The second question concerns the means of the possible adjustment. In utterance-medial position sibilants of two words can become adjacent to each other, like in *Das Salz* ([d̥asʰalts] the salt) and *Das Schild* ([d̥asʃiltʰ] the sign). Unlike in *Das Salz*, the sequence in *Das Schild* not only allows varying the sibilant pitch by changing the spectral energy distributions and hence the noise qualities of the individual sounds. In addition, the sibilant pitch in *Das Schild* can be varied by changing the degree of the regressive alveolar-to-postalveolar assimilation (we refrain using from the conceptual label ‘place assimilation’ here, as the assimilation affects more the shape of the tongue than the place of the constriction it creates). The sibilant pitches conveyed by the postalveolar [ʃ] are inherently considerably lower than those of [s]. So, stronger /s/-to-[ʃ] assimilations in the time and/or frequency domains will result in overall decreasing sibilant pitch impressions. So far, it has only been shown that the degree of assimilation across word boundaries is *indirectly* related to prosodic structure and varies, for example, as a function of speech rate, emphasis, or the level in the prosodic hierarchy at which the assimilation occurs [3, 4]. However, is the degree of assimilation also *directly* linked with prosodic structure as a constituting component of the intonation contour?

In order to address these questions, we measured the noise characteristics of utterance-medial /sz/ and /sʃ/ sequences in German statements and questions with regard to sequence durations and spectral centre-of-gravity (CoG) means and ranges. Since German has progressive voice assimilation, /sz/ always resulted in [s̥z̥] or [ss] sequences so that voicing did not interfere

with the CoG measurements. In addition to the high-pitched statement and low-pitched question conditions, we also included two expressiveness conditions, i.e. neutral vs. emphatic, in the experiment. Provided that the degree of assimilation in /sʃ/ will vary in accord with the intonation contexts, the variations will interact with the expressiveness conditions in a way that sheds new light on existing interpretations of segmental exponents of emphasis.

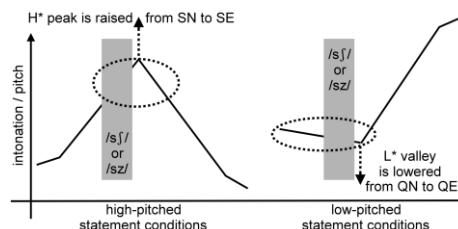
## 2. METHOD

The primary and secondary research questions were investigated using an existing corpus of Northern Standard German read speech. It contained 12 target pairs of function words and nouns like (a) *aus Schweden* ([aʊs ˈʏe:dʰn] from Sweden), (b) *als Sanger* ([alts zenɐ] as a singer), (c) *bis Sachsen* ([bɪs zaksʰn] to Saxony), (d) *als Spender* ([alts ʃˈpendɐ] as a donor). Word pairs like (a)-(b) constituted the non-assimilatory sibilant-sequence condition /sz/; word pairs like (c)-(d) created the assimilation condition /sʃ/ across word boundaries. The 12 target word pairs occurred as direct-object phrases in simple syntactically declarative (i.e. SVO) sentences, which in turn were embedded in longer dialogue frames on everyday topics like traveling, lectures, and friendship. These frames (and the corresponding punctuations) formed semantic-pragmatic contexts that supported the reading of the target sentences as neutral statements (SN), emphatic statements (SE), neutral questions (QN), and emphatic questions (QE).

In all 2x2 sentence mode and emphasis conditions the nouns received a pitch accent. These were H\* accents in the statements and L\* accents in the (also syntactically declarative) yes/no questions (cf. Fig.1). The sibilant sequences interrupted F0 close to the vowel-internal H\* peaks and L\* valleys and hence occurred in clearly distinct high-pitch vs. low-pitch intonation conditions. The emphasis contexts elicited vehement disagreement (with contrastive focus on the target nouns) in the statements and disbelieving astonishment (on the target nouns) in the questions. The emphases raised the H\* peaks of the statements and lowered the L\* valleys (including the preceding shallow F0 falls) of the questions. Thus, as is displayed in Figure 1, the high vs. low pitch contrast of questions and statements is enhanced under emphasis, cf. [7].

So, if the sibilant-sequence qualities reflect the surrounding F0 level, then the sibilant pitches of the sequences should decrease across the 4 context conditions in the order SE > SN > QN > QE.

**Figure 1:** Schematic representation of the sibilant-sequence locations in the F0 patterns of the elicited statements (SN, SE, left) and questions (QN, QE, right).



The sibilant pitches and the degrees of assimilation in the /sʃ/ sequences were estimated by means of CoG measurements within the manually annotated sibilant-sequences boundaries. Across each sibilant sequence CoG measurements were taken with *Praat* at intervals of 10 ms and within a frequency range of 2-12 kHz. The frequency range covers the main spectral characteristics of [s] and [ʃ] and excludes potential F0 residuals as well as high-frequency noise of the recording. With reference to [6] and [9], we can assume that CoG is a suitable estimate of perceived sibilant pitch. The number of CoG measurements varied with the overall duration of the sibilant sequence. Based on the variable number of measurements, a mean CoG and a CoG range were calculated for each sibilant sequence.

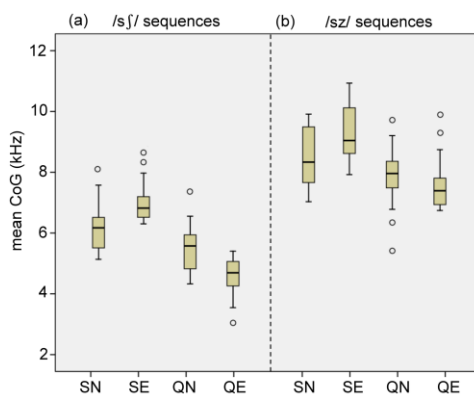
The dialogues were read by four pairs of female speakers (i.e. by 8 subjects in total). They were between 20-30 years old, grew up and lived in Northern Germany. The recordings were made digitally (96kHz, 32bit) in a sound-treated booth at the Institute of Phonetics (IPDS) of the University of Kiel. In order to create an informal, communicative situation in the lab the dialogue partners knew each other and the experimenters for a long time. They were given 30 min to practise the dialogues prior to the recording, and they were instructed to produce the dialogues as naturally and casually as possible in accord with the dialogues' everyday topics. All dialogues were produced twice in differently randomized orders with reversed roles of the speakers (for details cf. [7]).

## 3. RESULTS

It was decided to exclude one speaker from the data analysis, as she clearly overdid the instruction

to produce the texts in an everyday speaking style. Therefore the results summarized in Figures 2(a)-(b) are based on 7 speakers. As every speaker produced the sibilant-sequence conditions SN, QN, SE, and QE three times across 12 dialogues, the 2x4 box plots represent 21 productions each. It can be assumed that the displayed spectral sibilant-sequence characteristics, i.e. the mean CoGs, are positively correlated with sibilant pitch. The corresponding F0 contexts were, apart from some variability in alignment and scaling of the F0 movements, all realized as shown in Figure 1.

**Figure 2:** Mean CoGs for the utterance-medial /sʃ/ and /sz/ sequences, averaged in each of the four conditions SN, SE, QN, and QE across the 21 productions of the 7 female German speakers.



As regards the /sʃ/ sequences in Figure 2(a), the box plots show a substantial difference between the two question (Q) and statement (S) conditions. Averaged across the sibilant sequences the CoGs were clearly (i.e. about 1-2kHz) higher in the statement than in the question conditions. Moreover, this question-statement difference was greater for the emphatic than for the neutral sentences. The /sz/ sequences illustrated in Figure 2(b) yielded a similar results pattern, though less pronounced. Supplementary to the descriptive analysis a univariate ANOVA was done. It returned highly significant main effects for two of the three fixed factors, viz. sentence mode (question vs. statement;  $F_{[1,160]}=177.286$ ;  $p<0.001$ ;  $\eta_p^2=0.526$ ) and sibilant sequence (/sz/ vs. /sʃ/;  $F_{[1,160]}=474.719$ ;  $p<0.001$ ;  $\eta_p^2=0.748$ ). The third fixed factor, expressiveness (neutral vs. emphatic), had no separate significant effect on the mean CoGs. However, there was a highly significant interaction between expressive-ness and sentence mode ( $F_{[1,160]}=23.155$ ;  $p<0.001$ ;  $\eta_p^2=0.126$ ), as well as a marginally significant interaction between all three factors ( $F_{[1,160]}=4.427$ ,  $p=0.037$ ;

$\eta_p^2=0.027$ ). The interactions are due to the important facts that (a) the mean CoGs increase from neutral to emphatic sentences in statements and decrease from neutral to emphatic sentences in questions, and that (b) the CoG changes were greater for /sʃ/ than for /sz/.

An additional, analogously structured ANOVA was calculated for the sibilant-sequence durations. It was to make sure that the differences in mean CoGs reflect assimilation rather than elision processes. Crucially, the analysis showed that the /sz/ sequences were not longer than the /sʃ/ sequences. Instead, they were slightly, but significantly shorter (in ms:  $x_{/sz/}=106$ ;  $sd_{/sz/}=20$ ;  $x_{/sʃ/}=124$ ;  $sd_{/sʃ/}=19$ ;  $F_{[1,160]}=42.332$ ;  $p<0.001$ ;  $\eta_p^2=0.209$ ). This outcome is in line with the interpretation of the mean-CoG variation of /sʃ/ in terms of assimilation. The other two factors sentence mode and expressiveness also yielded significant main effects. In the case of sentence mode, both the /sz/ and /sʃ/ sequences were shorter in the question than in the statement sentences (in ms:  $x_S=119$ ;  $sd_S=21$ ;  $x_Q=111$ ;  $sd_Q=22$ ;  $F_{[1,160]}=8.441$ ;  $p=0.004$ ;  $\eta_p^2=0.050$ ). The neutral and emphatic sentences differed in that the latter showed almost 20 ms longer sibilant sequences (in ms:  $x_N=105$ ;  $sd_N=17$ ;  $x_E=124$ ;  $sd_E=21$ ;  $F_{[1,160]}=48.774$ ;  $p<0.001$ ;  $\eta_p^2=0.234$ ). None of the interactions were significant.

Mean CoG on its own is not sufficient to estimate the degree of assimilation in the /sʃ/ sequences. For example, it is possible that the spectral differences of separate /s/ and /ʃ/ segments were maintained in all productions, so that the mean CoG differences only reflect a temporal shift of the boundary between the sibilants (in favour of the postalveolar segment). Thus the mean CoG of each sequence was complemented by the CoG range, i.e.  $CoG_{max} - CoG_{min}$ . The CoG ranges shed light on the degree of assimilation in the spectral dimension. Smaller CoG ranges mean smaller spectral changes across the sibilant sequences, which in turn indicates a stronger spectral assimilation of the /s/ and /ʃ/ elements (towards intermediate or overall more [ʃ]-like sibilant qualities).

On this basis, we counted for each speaker and condition, how many CoG ranges of the /sʃ/ sequences were smaller than average CoG range of all /sz/ sequences plus one standard deviation. CoG ranges of /sʃ/ below the /sz/ threshold mean that the corresponding alveolar-postalveolar sequences

were realized spectrally as homogeneous as the purely alveolar sequences. This can be interpreted as an (almost) complete spectral /s/-to-[ʃ] (and hence /sʃ/-to-[ʃʃ]) assimilation. The count resulted in 18 instances for the QE condition, followed by 15, 12, and 4 instances in the QN, SN, or SE conditions, respectively. The order of these frequencies matches well with the box plots of the mean CoGs in Figures 2(a)-(b). An additional chi-square test showed that the frequencies differ significantly between the 2x2 sentence mode and expressiveness conditions ( $\chi^2=3.802$ ;  $df=1$ ;  $p=0.05$ ), mainly due to the  $\chi^2$  contributions of the emphatic conditions.

#### 4. CONCLUSIONS

The acoustic analysis of the 84 sibilant sequences showed for /sz/ that the mean CoGs varied systematically and in parallel with the F0 contexts of the statement and question conditions (SN, SE vs. QN, QE). That is, when preceding the high pitch peaks in the statement sentences, the /sz/ sequences were produced with higher mean CoGs than in the question sentences, where the /sz/ sequences preceded low pitch valleys. Supported by clear initial evidence from informal perception experiments (cf. also [http://www.linguistik.uni-kiel.de/Niebuhr\\_index.html/segminto.zip](http://www.linguistik.uni-kiel.de/Niebuhr_index.html/segminto.zip)) and with reference to [6, 9] it may be claimed that the mean CoG differences given in Figure 2 manifest themselves as clearly distinct sibilant pitches. So, as regards the primary research question, our study provided strong initial evidence that the adjustment of the pitch impression conveyed by voiceless fricatives (sibilants) to the F0-related pitch of the intonation context not just occurs utterance-finally, but also utterance-medially. The traditional view in which voiceless sound segments are troublemakers that interrupt pitch and interfere with the coding of intonational meaning, does not hold in a strict, general sense.

With regard to sibilant-pitch adjustment, the /sʃ/ sequences showed a qualitatively identical, but even stronger noise quality variation (in terms of mean CoGs) than the /sz/ sequences. At the same time, the /sʃ/ sequences were on average not shorter than the /sz/ sequences, which rules out that the variation found for /sʃ/ is merely due to /s/ reduction or elision. Furthermore, in the statement conditions the CoG ranges were by majority clearly greater than the /sz/ ranges. That is, compared with the small CoG ranges of the

consistently alveolar /sz/ sequences, the greater /sʃ/ ranges were reflect clear spectral transitions from alveolar to postalveolar noise qualities. However, in the question conditions the CoG ranges of /sʃ/ were almost all similar to the /sz/ ranges, i.e. the /sʃ/ productions approximated [ʃʃ]. Altogether, this allows the conclusion that, with respect to the secondary research question, the utterance-medial adjustment of sibilant pitch involves the degree of regressive /s/-to-[ʃ] assimilation. The sibilant pitch induced by /sʃ/ is raised/lowered by weakening/intensifying the /s/-to-[ʃ] assimilation. This direct link between intonation and assimilation strongly violates the segment-prosody dichotomy.

The CoG and concomitant sibilant-pitch changes under neutral and emphatic conditions followed the direction of the F0 scaling, which supports the conclusion that the changes relate to the intonation context and are not a by-product of the sentence mode. Finally, the fact that the degree of assimilation is lower in emphatic (i.e. narrow focus) statements was so far solely explained by reduced speech rate due to local hyperarticulation. This is inconsistent with our data, as the speech rate decreased for both emphatic statements and questions. The possibility that the lower degree of assimilation under emphasis could account for a raised F0 context has not been considered so far, which suggests revisiting the findings of previous studies.

#### 5. REFERENCES

- [1] Grabe, E. 1998. Pitch accent realisation in English and German. *Journal of Phonetics* 26, 129-144.
- [2] Kohler, K.J., Niebuhr, O. 2011. On the role of articulatory prosodies in German message decoding. *Phonetica* 68, 1-31.
- [3] Kuzla, C. 2009. Prosodic structure in speech production and perception. *MPI Series in Psycholinguistics* 52.
- [4] Mücke, D., Grice, M., Kirst, R. 2008. Prosodic and lexical effects on German place assimilation. *Proc. ISSP 2008* Strasbourg, France, 225-228.
- [5] Niebuhr, O. 2008. Coding of intonational meanings beyond F0: Evidence from utterance-final /t/ aspiration in German. *J. Acoust. Soc. Am.* 142, 1252-1263.
- [6] Niebuhr, O. 2009. Intonation segments and segmental intonations. *Proc. Interspeech* Brighton, UK, 2435-2438.
- [7] Niebuhr, O., Bergherr, J., Huth, S., Lill, C., Neuschulz, J. 2011. Intonationsfragen hinterfragt. To appear in *ZDL*.
- [8] Niebuhr, O., Kohler, K.J. 2011. Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, doi:10.1016/j.wocn.2010.12.003.
- [9] Traunmüller, H. 1987. Some aspects of the sound of speech sounds. In Schouten, M.E.H. (ed.), *The Psychophysics of Speech Perception*. Dordrecht: Martinus Nijhoff, 293-305.