Prosodic Structure Affects the Production and Perception of Voice-Assimilated German Fricatives

Claudia Kuzla, Mirjam Ernestus & Holger Mitterer

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
{Claudia.Kuzla; Mirjam.Ernestus; Holger.Mitterer}@mpi.nl

Abstract

Prosodic structure has long been known to constrain phonological processes [1]. More recently, it has also been recognized as a source of fine-grained phonetic variation of speech sounds. In particular, segments in domain-initial position undergo prosodic strengthening [2, 3], which also implies more resistance to coarticulation in higher prosodic domains [5]. The present study investigates the combined effects of prosodic strengthening and assimilatory devoicing on word-initial fricatives in German, the functional implication of both processes for cues to the fortis-lenis contrast, and the influence of prosodic structure on listeners' compensation for assimilation. Results indicate that 1. Prosodic structure modulates duration and the degree of assimilatory devoicing, 2. Phonological contrasts are maintained by speakers, but differ in phonetic detail across prosodic domains, and 3. Compensation for assimilation in perception is moderated by prosodic structure and lexical constraints.

1. Introduction

Prosodic structure refers to the organization of spoken language into hierarchically embedded domains: Syllables form words, which are grouped into phrases, and utterances may consist of several phrases. While the number and the phonetic correlates of prosodic domains may differ across languages, the influence of prosodic structure on both phonological processes [1] and on fine-grained phonetic details [2, 3, 4] has been demonstrated for a substantial number of languages. In French, for instance, *liaison* (i.e., the realization of an underlying word-final consonant, which is not present in the citation form of a word) occurs only if the following vowel belongs to the same phrase. Within the framework of Prosodic Phonology, the scope of such phonological processes defines prosodic domains. Another line of research has shown that different-sized prosodic boundaries, as defined mainly by intonational criteria, also affect the phonetic details of boundary-adjacent segments, resulting in 'final lengthening' and 'initial strengthening'. The focus of the present study is on initial strengthening, which implies increase in duration and in spatial articulatory expansion. As a consequence, segments are also less coarticulated across larger prosodic boundaries [5].

The present study investigates the effects of prosodic structure on the production and perception of the word-initial German fricatives /f, v, z/ [cf. 6]. The phonological contrast between /f, s/ (fortis) on the one hand and /v, z/ (lenis) on the other hand is mainly cued by two acoustic cues: duration and glottal vibration. In our production studies, we investigated how both cues are affected by initial strengthening: Our first hypothesis stated that the duration of initial fricatives would increase with the size of the prosodic domain. Second, the lenis fricatives /v, z/ can be devoiced if they follow voiceless obstruents, such as /t/. This assimilatory devoicing has been reported to be complete for /z/, which does not contrast with /s/ in word-initial position (e.g., /hat zant/ 'has sand' -> [hatsant]), but to be incomplete for /v/ (/hat vasɐ/ 'has water' -> [hatv̥asɐ]), which is assumed to remain distinguishable from /f/ [7, 8]. We hypothesized that the degree of assimilatory devoicing was not only conditioned by such phonotactic or lexical constraints, but also by the size of the prosodic boundary (Experiment 1).

The partial devoicing could make the recognition of words such as /vɛldɐ/ 'forests' harder for listeners, since there may exist competing words beginning with /f/, such as /fɛldɐ/ 'fields'. Hence, we investigated in three perceptual experiments how listeners deal with this variation in different prosodic domains. Perceptual studies on other types of assimilation have shown that listeners compensate for assimilation, that is, that they adjust their phonemic categories to viable versus non-viable assimilation contexts [9, 10]. These studies, however, have not yet taken prosodic constraints into account, despite the evidence that prosodic structure influences assimilatory processes [1]. We therefore investigated compensation for the progressive assimilatory devoicing of German fricatives with respect to the question whether prosodic structure modulated the expected compensation (Experiments 2, 3). Finally, we addressed the question whether the lexico-functional difference between the /f-v/-contrast versus the /s-z/-contrast has an effect on perception (Experiment 4).

2. Experiment 1: Production of /f, v, z/

Eight native speakers of Northern Standard German read various sentence types[1] as illustrated by the examples given in Table 1. The sentence types were constructed to elicit different prosodic boundaries before the target words. The targets words, starting with the three fricatives /f/, /v/, and /z/, were: Felder 'fields', Wälder 'forests', and Senken 'hollows'. The preceding context was varied to be /t/ in hat 'has' (=assimilation context), and /ə/ in hatte 'had' (=non-assimilation context). Speakers were instructed to produce a contrastive phrasal accent on another word in the utterance

[1] The data in the present study overlap partially with data reported earlier [6]; however, the prosodic and statistical analyses were elaborated, leading to new results.

(indicated in boldface in Table 1), leading ideally to deaccentuation of the target words. Each speaker produced four repetitions of each sentence. Recordings were segmented and prosodically annotated by two phonetically trained native speakers of German. Utterances containing a pause before, or an accent on the target word were excluded from analyses, since both a pause and accentuation may influence the voice assimilation and the duration of the targets. The remaining 601 utterances were grouped into two prosodic categories, Phrase and Word. A Phrase boundary was defined by the presence of a pitch movement associated with a 'boundary tone', a Word boundary by the absence of a melodic break. Posthoc examination of the preboundary syllable durations supported this classification, showing a significantly different lengthening pattern of Phrase > Word.

• Inducing Phrase Boundary, Assimilation Context
Weil sie vorhat, Felder und **Wiesen** zu malen, ...
Because she plans, fields and meadows to draw, ...
*'Since she wants to draw fields and **meadows**, ...'*
• Inducing Word Boundary, Non-assimilation Context
Anna hatte Felder und Wiesen gemalt.
Anna had fields and meadows drawn
'***Anna** had drawn hollows and hills.'*

Table 1: *Examples of sentence types for the sequences* /t(ə)#f/.

We measured the duration and the period of glottal vibration of the target fricatives. First, we investigated domain-initial strengthening separately from the potential prosodic influence on voice assimilation. We modeled the durations of /f/ as a function of Prosodic Boundary and Context, and of /v, z/ in the non-assimilation context as a function of Prosodic Boundary by fitting multi-level linear mixed models with Speaker as a random variable.

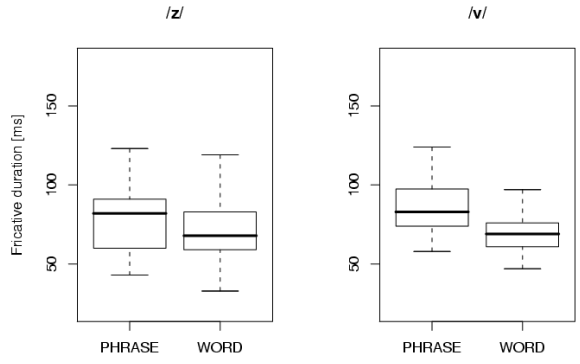

Figure 1: *Distribution of durations of /z/ and /v/ in /ə/-context as a function of prosodic boundary*

The analyses showed that all fricatives were significantly longer after a Phrase boundary than after a Word boundary, as illustrated in Figure 1 for /v/ and /z/. The effect on the duration of /f/ was similar and did not differ between the two contexts. These results confirm the previously reported evidence for domain-initial strengthening in German [6].

Second, we studied the influence of prosodic structure on the degree of assimilatory devoicing for /v, z/ in assimilation context (/t/). We normalized the period of glottal vibration for the duration of the fricative, by analyzing the percentage of the fricative produced with glottal vibration, instead of the plain duration of glottal vibration. A linear mixed model with Prosodic Boundary and Fricative as fixed effects and Speaker as random effect showed a significant difference in the degree of devoicing between the two fricatives: /v/ is devoiced to a lesser degree than /z/. Apart from an articulatory-aerodynamic explanation based on the difference in oral cavity size, which makes continuous vocal fold vibration harder to maintain in alveolars than in labiodentals, the lexico-functional load of the /f-v/ contrast may have played a role. In contrast to the /f-v/ distinction, the phonemic difference between /s/ and /z/ is neutralized in word-initial position in German, where /s/ is phonotactically illegal. In other words, the assimilation in /hat zɛŋkən/ 'has hollows' does not produce lexical competition, since */sɛŋkən/ is not an existing word.

More importantly, both fricatives were significantly more assimilated across Word boundaries than across Phrase boundaries (Figure 2). This finding shows that prosodic structure does not only constrain the occurrence of assimilation processes, but may also modulate their degree.

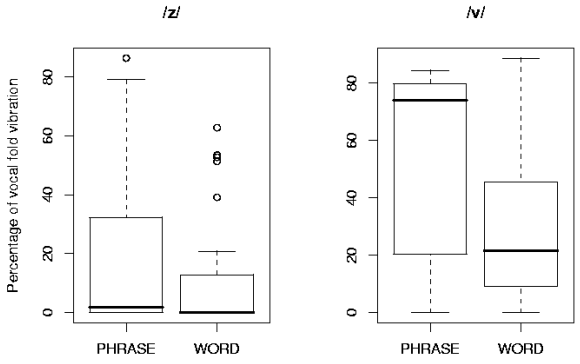

Figure 2: *Distribution of percentage of vocal fold vibration during /z/ and /v/ in /t/-context as a function of prosodic boundary*

Finally, we investigated whether assimilatory devoicing had any effects independent of domain-initial lengthening on the duration of assimilated fricatives. Since assimilatory devoicing makes (voiced) lenis fricatives more fortis-like (i.e., unvoiced), it might also affect duration as another cue. If so, more assimilation at smaller boundaries would lead to an increase in duration, which would counteract the domain-initial lengthening. On the other hand, the overlap and reduction of articulatory gestures which may underlie assimilation might also shorten more assimilated fricatives. We compared the durations of /v, z/ across contexts and examined the correlation of percentage of glottal vibration

with duration within prosodic conditions. However, we did not find any evidence for assimilatory devoicing influencing segment duration in either direction.

Our results have provided evidence for prosodic strengthening in German and have shown that this initial strengthening also implies greater resistance to progressive assimilatory devoicing, which appeared to be a gradient, prosodically-conditioned process for both /v/ and /z/. With respect to the fortis-lenis distinction which may be challenged by assimilatory devoicing, prosodic structure affects the two major cues into opposite directions: At higher boundaries, lenis fricatives are longer, thus more fortis-like, but also less devoiced, thus more lenis-like. This suggests that prosodic strengthening minimizes the assimilation effect on the categorical identity of lenis fricatives, resulting in a stable dispersion of phonemic categories in all contexts.

Given that in production, we found less glottal vibration after /t/ than after /ə/, especially after a Word boundary, we hypothesized that in perception, less glottal vibration would be required for a /v/-percept in a viable assimilation context than in a non-viable context (compensation for assimilation); and that within /t/-context, less glottal vibration would be necessary across smaller boundaries (compensation for prosodic structure). These predictions were tested in two phoneme categorization experiments (Experiments 2, 3).

3. Experiment 2: Perception of devoiced /v/

In a first experiment, we tested the perceptual effects of amount of glottal vibration during /v/. The effect of the amount of glottal vibration for the perception as fortis or lenis was investigated in four conditions with varying segmental context and prosodic boundary between context and target fricative: Assimilation context + Word boundary, Non-assimilation context + Word boundary, Assimilation context + Phrase boundary, Non-assimilation context + Phrase boundary. We generated two voicing continua based on two natural tokens of /v/, a fully voiced and a completely devoiced token as uttered by a female speaker who was reading part of the materials in Experiment 1. The continuum sounds generated with PRAAT interpolated between these endpoint by steps of single glottal cycles. The durations of the test sounds differed for the two continua since they were chosen to match the two experimental prosodic conditions, Phrase and Word, and showed typical values for the speaker (70 ms versus 50 ms). As test sounds, we selected 7 steps from each continuum which were matched in percentage of glottal vibration, equidistant within the range between 15% and 85%. The test sounds were spliced into carrier sentences similar to those in Table 1 as uttered by the same female speaker, where prosodic context was varied between Phrase and Word as defined above, and segmental context was varied between /t/ (assimilation context) and /ə/ (non-assimilation context). 17 native listeners identified the target words in utterance contexts as 'Felder' or 'Wälder' in a fully crossed design (2 Boundaries x 2 Contexts x 7 Steps) with 10 repetitions for each stimulus. Figure 3 illustrates the averaged /v/-identification scores for the four prosodic+segmental context conditions.

The data were subjected to ANOVA and posthoc comparisons by t-tests. In both prosodic conditions, we found a significant context effect: more /v/-responses were given in assimilation context (/t/) than in non-assimilation context (/ə/) confirming the compensation for assimilation hypothesis. Interpretation of the prosodic boundary effect was complicated, since, contrary to our expectations, it did not reach significance within /t/-context, but was present in /ə/-context. Overall, there was a strong bias towards /v/-responses in almost all experimental conditions (except for the maximally devoiced endpoint in the Phrase-Schwa condition), which may be due to the short duration of the fricative, a cue for lenis. Hence, we may attribute these results to a ceiling effect. Therefore, we ran a second identification experiment with a full /f-v/ continuum.

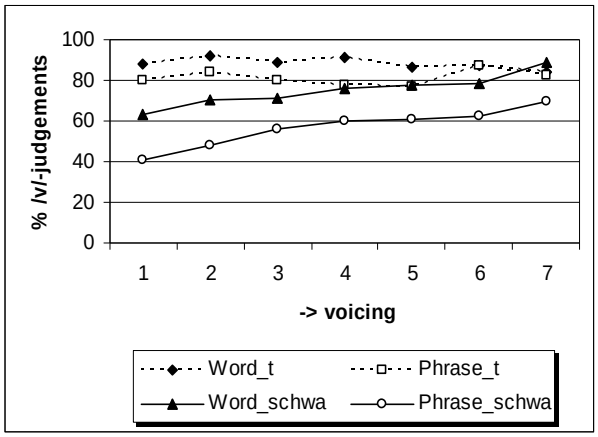

Figure 3: *Percentages of /v/-judgments across duration-normalized voicing continua in four prosodic+segmental context conditions*

4. Experiment 3: Perception of /f-v/

In this Experiment, we used a similar procedure as in Experiment 2. However, we generated a single continuum between clear /f/- and /v/-endpoints, that is, test sounds showed increasingly longer durations and up to zero cycles of glottal vibration towards the /f/-endpoint [120 ms], and continuous glottal vibration at the /v/-endpoint [70 ms]. The interpolation of duration and number of glottal cycles yielded 20 continuum steps, which were all used in the experiment. As in Experiment 2, the crossing of two prosodic and two segmental factors resulted in four conditions. Because of the larger number of steps we presented only four repetitions. 20 native listeners again identified target words in full utterance contexts as 'Felder' or 'Wälder'. The mean percentages of /v/-responses are plotted for the four prosodic+segmental conditions in Figure 4.

Analysis of variance and posthoc comparisons showed the significance of viable versus non-viable assimilation context in both prosodic conditions, except for steps close to the endpoints in most conditions. The previous finding of compensation for coarticulation has thus been replicated. In addition, we found compensation for prosodic structure in the /t/-context, in that more /v/-responses were obtained in the Word condition than in the Phrase condition. This effect appeared also in a second analysis: we fitted logistic regression models to the categorization functions of each participant and performed an ANOVA over the aggregated calculated 50% points (phoneme boundaries). The boundary between /f/ and /v/ shifted to a more devoiced step on the continuum in the Word condition. Note that this category shift

cannot be attributed to the fact that because of the interpolation procedure in this experiment, a more devoiced sound also had a longer duration, since longer duration would have served as a stronger cue to fortis (/f/) and would have counteracted the compensation for prosodically-conditioned assimilation.

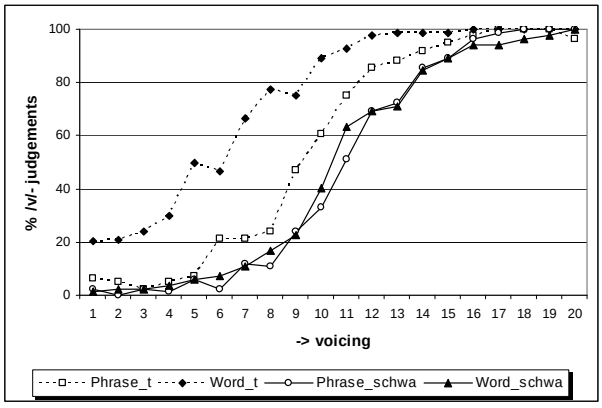

Figure 4: *Percentages of /v/-judgments across /f-v/-continuum in four prosodic+segmental context conditions*

5. Experiment 4: Perception of /s-z/

In a fourth experiment, we addressed the question whether the difference between the fricatives observed in Experiment 1—more assimilatory devoicing for /z/ than for /v/ (Figure 2)—also had consequences in perception. We generated a 20-step /s-z/ continuum in the same way as described in Experiment 3 for /f-v/, and ran the same procedure with 20 new participants. This time, they had to decide whether the stimulus they heard was the existing target word /z/*enken* 'hollows', or a "mispronunciation" resulting in the non-word **senken*. The results replicated the context effect, that is, listeners compensated for assimilation even in case of a non-phonemic contrast. Interestingly, we did not find any difference between the prosodic conditions in the assimilation context (/t/). In contrast to Experiment 3, listeners did not exploit the subtle phonetic differences induced by prosodic structure. Prosodic structure seems to be exploited more readily if the recognition of existing words is facilitated (as in Experiment 3).

6. General Discussion

The present study addressed the role of prosodic structure in the production and perception of German word-initial fricatives. We focused on the process of assimilatory devoicing of lenis fricatives across different-sized prosodic boundaries. This sandhi process might have caused a potential problem for listeners in the case of /v/, since there exist competing words beginning with /f/. Moreover, other cues to the fortis-lenis distinction are also affected by domain-initial prosodic strengthening.

The production study reported in Experiment 1 revealed that there is indeed prosodic strengthening in German fricatives, both in terms of phonetic expansion and of resistance to assimilation. Prosodic strengthening of lenis fricatives in assimilation environments affects two major cues to the fortis-lenis distinction in a way which keeps phonemic contrasts stable: At smaller boundaries, lenis fricatives are more devoiced, but also shorter than at larger boundaries. Furthermore, the assimilation process seems to be constrained by the lexico-functional load of the fricative identity (Experiment 1).

Our perception studies have shown that listeners compensate for assimilatory devoicing in viable contexts, and that this compensation is also moderated by prosodic structure. However, compensation for the influence of prosodic structure on the amount of glottal vibration only becomes visible in the phoneme identification paradigm if other cues, in particular duration, do not override them (Compare Experiments 2 and 3). Moreover, the results of Experiment 4 suggest that listeners only take prosodic structure into account if it is relevant in the comprehension of existing words. It cannot, however, be ruled out that the difference between /f-v/ and /s-z/ is not drive by the lexico-functional load, but rather by articulatory differences. Further research with methods which provide a better temporal resolution of activation processes in the mental lexicon is desirable.

In summary, this study sketches a complex interaction of prosody, sandhi processes, and phonotactic constraints, maintaining clear acoustic cues to prosodic structure as well as to the fortis-lenis distinction for the listener.

7. References

[1] Nespor, M.; Vogel, I., 1986. *Prosodic Phonology*. Dordrecht: Foris.

[2] Keating, P. A.; Cho, T., Fougeron, C.; Hsu, C.-S. ,2003. Domain-initial articulatory strengthening in four languages. In *Papers in Laboratory Phonology VI* , J. Local; R. Ogden; R. Temple (Eds.). Cambridge: Cambridge University Press, 145-163.

[3] Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetic*, 29, 109-135.

[4] Wightman, C. W.; Shattuck-Hufnagel, S.; Ostendorf, M.; Price, P. J., 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91, 1707-1717.

[5] Cho, T., 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32, 141-176

[6] Kuzla, C., 2003. Prosodically-conditioned variation in the realization of domain-final stops and voicing assimilation of domain-initial fricatives in German. *Proceedings of the 15th ICPhS, Barcelona/Spain*, 2829-2832.

[7] Kohler, K. J., 1995. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

[8] Jessen, M., 1998. *Phonetics and Phonology of tense and lax obstruents in German.* Amsterdam: John Benjamins.

[9] Gow, D., 2003. Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics* 65, 575-590.

[10] Mitterer, H.; Blomert, L., 2003. Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics* 65, 956-969.