

Prosodic strengthening of German fricatives in duration and assimilatory devoicing

Claudia Kuzla^{a,*}, Taehong Cho^{a,b}, Mirjam Ernestus^{a,c}

^a*Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands*

^b*Division of English Language and Literature, Hanyang University, Seoul, Republic of Korea*

^c*Radboud University Nijmegen, The Netherlands*

Received 10 February 2006; received in revised form 27 September 2006; accepted 6 November 2006

Abstract

This study addressed prosodic effects on the duration of and amount of glottal vibration in German word-initial fricatives /f, v, z/ in assimilatory and non-assimilatory devoicing contexts. Fricatives following /ə/ (non-assimilation context) were longer and were produced with less glottal vibration after higher prosodic boundaries, reflecting domain-initial prosodic strengthening. After /t/ (assimilation context), lenis fricatives (/v, z/) were produced with less glottal vibration than after /ə/, due to assimilatory devoicing. This devoicing was especially strong across lower prosodic boundaries, showing the influence of prosodic structure on sandhi processes. Reduction in glottal vibration made lenis fricatives more fortis-like (/f, s/). Importantly, fricative duration, another major cue to the fortis-lenis distinction, was affected by initial lengthening, but not by assimilation. Hence, at smaller boundaries, fricatives were more devoiced (more fortis-like), but also shorter (more lenis-like). As a consequence, the fortis and lenis fricatives remained acoustically distinct in all prosodic and segmental contexts. Overall, /z/ was devoiced to a greater extent than /v/. Since /z/ does not have a fortis counterpart in word-initial position, these findings suggest that phonotactic restrictions constrain phonetic processes. The present study illuminates a complex interaction of prosody, sandhi processes, and phonotactics, yielding systematic phonetic cues to prosodic structure and phonological distinctions.

© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

The fine phonetic details of segment realizations are determined both by prosodic structure (e.g., Byrd, Kaun, Narayanan, & Saltzman, 2000; Cho, 2004; Fougeron, 2001; Keating, Cho, Fougeron, & Hsu, 2003; Pierrehumbert & Talkin, 1992; Shattuck-Hufnagel & Turk, 1996; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992) and by sandhi processes such as coarticulation and assimilation (cf. Kühnert & Nolan, 1999). The variable phonetic patterns due to prosodic structure are often described under the rubric of ‘prosodic strengthening’ which can be defined as “temporal and/or spatial expansion of articulation due to accent and/or prosodic boundary” (Cho, 2005). Importantly, sandhi processes are themselves also constrained by prosodic structure: They are typically more robust and frequent across lower than higher prosodic boundaries

*Corresponding author. Tel.: +31 24 3521334; fax: +31 24 3521213.

E-mail address: Claudia.Kuzla@mpi.nl (C. Kuzla).

(cf. Nesp̄or & Vogel, 1986; Selkirk, 1986). In the present study, we examine the combined effects of prosodic strengthening and a sandhi process in a context where the two factors are expected to affect the acoustic realization of fricatives in opposite directions.

More specifically, this study examines prosodic strengthening in combination with assimilatory devoicing in German fricatives. Both processes may affect acoustic cues to the fortis-lenis distinction in fricatives (Fischer-Jørgensen, 1963; Jessen, 1998 for German; Cole & Cooper, 1975; Pirello, Blumstein, & Kurowski, 1997 for English). Initial strengthening may make fricatives more fortis-like in higher prosodic domains, but assimilatory devoicing may make them more fortis-like in lower prosodic domains. The combined effect of the two prosodically conditioned processes may depend on how exactly they influence the different acoustic cues to the fortis-lenis distinction. Furthermore, the combined effect may depend on language-specific lexical/phonotactic constraints (e.g., the /s/-/z/ contrast in German does not occur in word-initial position whereas the /f/-/v/ contrast does). We investigated the fine-grained acoustic realization of German fricatives in three different prosodic conditions and in assimilation and non-assimilation contexts.

1.1. Phonetic and phonological correlates of prosodic structure

Many studies in various languages, including Dutch, English, French, Korean, and Taiwanese, have shown that prosodic structure affects the fine acoustic details of segment realizations. A well-known acoustic correlate of prosodic structure is ‘Final Lengthening’: Domain-final syllables are longer than medial ones (Wightman et al., 1992). Another correlate is ‘Initial Strengthening’: Consonants show more linguo-palatal contact (Fougeron & Keating, 1997; Cho & Keating, 2001; Fougeron, 2001), stops show longer closures and longer Voice Onset Time (Pierrehumbert & Talkin, 1992; Jun, 1998; Keating et al., 2003), and vowels are more often glottalized (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996) and show greater resistance to vowel-to-vowel coarticulation (Cho, 2004) in domain-initial than in domain-medial position. Both Final Lengthening and Initial Strengthening are assumed to be generally cumulative in the vertical dimension of the prosodic hierarchy, that is, the higher the prosodic domain, the stronger the effects.

Initial Strengthening suggests that prosodic structure might also influence sandhi processes, such as assimilation: Sandhi effects on initial segments may be weaker at higher prosodic boundaries (cf. Cho, 2004 for similar coarticulatory effects). This hypothesis is in accordance with evidence collected within the framework of Prosodic Phonology, where prosodic constituents are explicitly defined as application domains of phonological processes (Nesp̄or & Vogel, 1986; Selkirk, 1986; Jun, 1998). An example is French *liaison*, that is, the realization of an underlying word-final consonant before a following vowel (see /t/ in 1a), whereas the consonant does not surface in the citation form of the word (1b) or before a following consonant (1c). *Liaison* only applies between words that belong to the same phonological phrase, but not across phrase boundaries (example 1d, taken from Nesp̄or & Vogel, 1986, p. 179):

- (a) vocalic context: *les visiteurs sont* [s̄z̄ t] *arrivés* ‘the visitors have arrived’,
- (b) citation form: *sont* [s̄z̄] ‘(they) are’,
- (c) consonantal context: *les visiteurs sont* [s̄z̄] *partis* ‘the visitors have left’,
- (d) phrase boundary: *Le garçon les aidait* [ɛdɛ] // *activement*. ‘The boy helped them actively’.

1.2. Assimilatory devoicing of German fricatives

In the present study, we addressed the role of prosodic structure in the realization of word-initial fricatives in German, and explicitly investigated the possible interaction of domain-initial prosodic strengthening and assimilatory devoicing. We focused on /v, z, f/ in word-initial position. The fricatives /v/ and /z/ are the only German lenis fricatives which occur word-initially in the native vocabulary. Importantly for our research, /z/ has been described to be realized as [s] after voiceless obstruents (Kohler, 1990, p. 79; Kohler, 1995, p. 160). For example, *sich* ‘her-/himself’ is pronounced as [zɪç] in isolation, but the sequence *hat sich* ‘has her-/himself’ as [hatsɪç]. However, the assimilation is sometimes incomplete (Jessen, 1998), that is, it is gradient rather than categorical. /v/ also is assimilated after voiceless obstruents, but always remains distinguishable from /f/ (Kohler, 1995).

This fricative assimilation process is generally known as Progressive Voice Assimilation. The term is based on the assumption that /v, z/ differ from /f, s/ in the feature [+voice]. However, it has been shown that this phonological contrast is cued by multiple phonetic features, of which glottal vibration is not necessarily the most important one. Therefore, the phonemic contrast is often referred to as a contrast between ‘lax’ or ‘lenis’ fricatives (/v, z/) on the one hand, and ‘tense’ or ‘fortis’ fricatives (/f, s/) on the other hand. In line with this terminology, and in order to reserve the term ‘voicing’ for vocal fold vibration, we refer to the Progressive Voice Assimilation as ‘Assimilatory Fortition’. ‘Assimilatory devoicing’, on the other hand, refers to context-dependent (possibly gradient) reduction of the period of vocal fold vibration only.

1.3. Phonetic correlates of the fortis-lenis contrast

One of the main phonetic correlates of the fortis-lenis opposition in fricatives is frication duration. Fortis fricatives are longer than lenis ones in German, as in other languages (e.g., Fischer-Jørgensen, 1963; Jessen, 1998 for German; Kissine, Van de Velde & van Hout, 2003 for Dutch; Cole & Cooper, 1975; Pirello et al., 1997; Stevens, Blumstein, Glicksman, Burton, & Kurowski, 1992 for English; Fischer-Jørgensen, 1963 for Danish). However, studies in languages other than German suggest that duration also cues other phonemic distinctions. Shortening of fortis fricatives may affect the perception of manner of articulation, shifting from fricative to affricate (Ferrero, Pelamatti, & Vaggies, 1979 for Italian), or the perception of place of articulation (Jongman, 1989 for English) rather than turning the fortis fricative into a lenis fricative.

Vocal fold vibration has been established as another major cue to the fortis-lenis distinction in fricatives (e.g., Fischer-Jørgensen, 1963 for German; Slis & Cohen, 1969a, b; van den Berg & Slis, 1985; Kissine et al., 2003 for Dutch, but see Jessen, 1998 for a description of [voice] as a feature different from fortis/lenis). In general, /v, z/ are produced with vocal fold vibration, whereas /f, s/ are not. However, /v, z/ can be devoiced to a variable degree (Jessen, 1998 for German; Haggard, 1978; Pirello et al., 1997; Stevens et al., 1992 for English). In non-assimilatory context, this partial devoicing can be explained on articulatory and aerodynamic grounds. In order to initiate vocal fold vibration at the beginning of an utterance, a critical difference of about 400 Pa (4000 dyne/cm²) must be created between subglottal and supraglottal air pressure (Westbury & Keating, 1986; Baer, 1975). Since subglottal pressure is built up linearly utterance-initially, this threshold is difficult to overcome right at the beginning of an utterance. Similarly, a sufficient transglottal pressure differential must be maintained for continuous voicing of speech sounds in medial positions. Due to the oral impedance in obstruents, oral pressure increases over time, and vocal fold vibration ceases unless compensatory articulatory strategies (e.g., lowering of the larynx) are used to maintain the transglottal pressure differential. In voiced fricatives, transglottal airflow and airflow through the oral constriction must therefore be carefully balanced to produce vocal fold vibration and frication noise at the same time (Stevens et al., 1992; Stevens, 1998). Failure to do so leads to partial devoicing. In an experimental acoustic study, Jessen (1998) found that German /v, z/ were typically partially devoiced, but most tokens were produced with more than 20 ms of vocal fold vibration, which differentiated them reliably from fortis fricatives.

1.4. Phonotactic asymmetry between /f-v/ and /s-z/ in German

In German, both /f, v/ and /s, z/ display the fortis-lenis contrast. In the standard variety spoken in northern regions of Germany, however, there is a distributional asymmetry between /z/ and /s/. While both /v/ and /f/ occur word-initially, of the pair /s, z/ only /z/ is phonotactically legal in that position (Kohler, 1995; Jessen, 1998, p. 177), except for some recent English loanwords such as *cent*, *city*, *sex*. Complete assimilation of /v/ could thus lead to neutralization of lexical contrasts, such as *Wälder* [vɛldɐ] ‘forests’ versus *Felder* [fɛldɐ] ‘fields’, whereas assimilation of /z/ does not create potential lexical ambiguities. This leads to the hypothesis that assimilatory fortition may be complete for /z/, but not for /v/. Also Jessen (1998, p. 188) has suggested that “[...] in a language or dialect with a higher functional load of the opposition between /s/ and /z/ in this position, the degree and proportion of devoicing of /z/ would be lower.”

1.5. The prosodic structure of German

There is no wide consensus about the number of prosodic levels in German (Wiese, 1996; Fox, 1993). Most researchers agree upon the existence of a domain similar to the Prosodic Word, which is commonly defined as the domain of stress assignment, of syllabification, and of phonotactic constraints (e.g., Wiese, 1996; Booij, 1995; Peperkamp, 1997). In addition, the Intonation Phrase is generally accepted as the unit of a complete intonation contour that contains one or more pitch accents. It is mainly characterized by final lengthening, optional surrounding pauses, and a pitch movement corresponding to a ‘boundary tone’ (e.g., Silverman et al., 1992 for English; Grice & Baumann, 2000 for German). In contrast, the existence of an intermediate phrase level is controversial in German (Féry, 1993; Grice & Baumann, 2000, Grabe, 1998; Uhmman, 1991).

In the present study, we distinguished three prosodic conditions. The word level was separated from the phrase level. Furthermore, phrases which were preceded by a pause were distinguished from those which were not, for the following reasons. First, de Pijper and Sanderman (1994, for Dutch) showed that the presence of a pause is a strong perceptual cue to a high prosodic boundary in read speech. When listeners rated perceptual boundary strength on a 1–10 scale, highest scores were found in conditions where tonal cues co-occurred with a pause. When only one of the acoustic cues was present, higher scores were obtained for a pause than for a melodic cue. Second, from an articulatory-aerodynamic perspective, a pause is likely to affect glottal vibration, one of the foci of the present study. With respect to assimilatory devoicing, initiation of vocal fold vibration after a pause is physiologically difficult (as we described above).

1.6. Hypotheses tested in the present study

The duration and the voicing of fricatives were investigated in the three prosodic conditions (word boundary, minor phrase boundary, major phrase boundary), and in two segmental contexts. We chose an assimilation context /t#_/, where /v/ and /z/ could be devoiced by Assimilatory Fortition, and a vocalic context /ə#_/, where they could not.

Concerning the influence of prosodic structure on the duration of initial segments and on Assimilatory Fortition, two hypotheses were considered. First, the duration of both fortis /f/ and lenis /v, z/ in the non-assimilation context /ə#_/ is expected to increase at the beginning of higher prosodic domains, as is the case in other languages (the Initial Lengthening Hypothesis). Second, for lenis /v, z/ in the assimilation context /t#_/, an increase of glottal vibration across larger prosodic boundaries is predicted, since the impact of the preboundary /t/ on the following /v, z/ is smaller across such boundaries (the Assimilatory Devoicing Hypothesis).

With respect to the duration of /v, z/ in the assimilation context, we considered the following three competing hypotheses regarding the interaction of prosodic strengthening and assimilation, which were tested one by one. The first hypothesis is that initial lengthening is the only predictor for fricative duration. If so, fricatives in both assimilation and non-assimilation contexts are longer at higher prosodic boundaries (the Initial Lengthening Hypothesis). The second hypothesis is based on the observation that fortis fricatives are in general longer than lenis fricatives. More assimilation and therefore more fortis-like fricatives are expected at lower boundaries. This would result in longer fricatives at lower boundaries (the Fortition Lengthening Hypothesis). The third hypothesis states that assimilated fricatives are shorter than unassimilated ones because the assimilation may result from gestural reduction and overlap (the Overlap Shortening Hypothesis; cf. Browman and Goldstein, 1992; Zsiga, 2000; Mitterer, 2003). Since less assimilation may be expected at higher boundaries, the Overlap Shortening Hypothesis predicts the same durational effect as the Initial Lengthening Hypothesis, that is, longer fricatives at higher boundaries.

A final issue that we addressed concerns the difference between the fricatives /v/ and /z/. The more fronted place of articulation of /v/ implies a relatively larger supraglottal cavity compared to /z/, which facilitates glottal vibration for /v/. This leads to the prediction that /v/ is produced with more vocal fold vibration than /z/ (Aerodynamics Hypothesis). In addition, since the phonological contrast is maintained word-initially for /f/-/v/, but not for /s/-/z/, the /f/-/v/ contrast can be said to have a higher functional load in German than the /s/-/z/ contrast, especially in word-initial position. Therefore, less assimilatory fortition devoicing for /v/ than

for /z/ may be expected (Phonotactic Constraint Hypothesis). Note that both the Aerodynamics Hypothesis and the Phonotactic Constraint Hypothesis predict that /v/ is produced with a larger amount of vocal fold vibration than /z/.

2. Method

German speakers read meaningful sentences in which the fricatives /f, v, z/ occurred in various prosodic and segmental environments. The recordings were prosodically categorized, and voicing and duration measurements of the fricative and the surrounding segments were taken.

2.1. Participants

Ten native speakers of Northern Standard German participated in the experiment, five females, and five males. They had all been brought up in the Northern German federal states Schleswig-Holstein and Niedersachsen, and did not show any other regional influences in their pronunciation, as judged by the first author during conversation and a short reading task. All speakers were undergraduate students at Kiel University, aged 20–25, 22 years on average, and were paid for their participation.

2.2. Speech materials

We examined word-initial /f/, /v/, and /z/ (/s/ does not occur word-initially in the native vocabulary). In order to obtain as natural speech data as possible under a maximally controlled experimental approach, we constructed meaningful sentences (see Table 1) in which various factors, such as phonetic context, lexical stress, and phrasal accentuation were kept constant.

The target fricatives /f, v, z/ were always followed by the mid-low vowel /ɛ/ in the disyllabic plural nouns *Felder* [fɛldɐ] ‘fields’, *Wälder* [vɛldɐ] ‘forests’, and *Senken* [zɛŋkən] ‘hollows’, with primary stress on the first syllable. The preceding context was the voiceless stop /t/ from *hat* ‘has’, or /ə/ from *hatte* ‘had’. The /t/-context was chosen as the triggering condition for assimilation, and the /ə/-context served as the non-assimilatory condition.

To elicit prosodic boundaries of different sizes between the target fricative and the preceding segment (e.g., *hat#Felder*), we created sentences with four different syntactic structures, referred to as ‘sentence types’ below. The sentence set is given in Table 1. Note that we do not claim any direct mapping between syntactic structure and higher-level prosodic structure. However, we expected these sentences to be produced with different prosodic boundaries. Later, the actual realizations were classified based on a prosodic analysis (see below). We expected large prosodic boundaries to occur between sentences, where orthography requires a period (Table 1, Sentence). Such a boundary would be equivalent to the “utterance” boundary in Nespors and Vogel (1986), but also to the Intonation Phrase (IP) boundary in Selkirk (1986) and Beckman and Pierrehumbert (1986). Relatively smaller phrase boundaries, if they exist, are likely to occur between items of a list, and between larger syntactic constituents, such as a superordinate clause’s verb and a following NP in an obligatory complement clause (Table 1, List and Complement, respectively). These boundaries would be roughly equivalent to the intermediate phrase (ip) boundary of Beckman and Pierrehumbert (1986) or the phonological phrase boundary of Selkirk (1986) and Nespors and Vogel (1986). A prosodic word boundary is likely to occur between words within smaller syntactic units, for instance, the object and the finite verb form in a verb phrase (Table 1, Word).

To control for a potential confounding effect of phrasal accent on the realization of the target fricatives, speakers were asked to place a contrastive accent on a non-target word in the utterance (as indicated in bold case in Table 1). This resulted in non-accentuation of the target words.

The full crossing of experimental factors yielded 24 conditions, containing one test sentence each: 3 fricatives (f, v, z) × 2 preceding contexts (/t/ vs. /ə/) × 4 sentence types (Word, List, Complement, Sentence). Each speaker produced five repetitions of all 24 test sentences. The sentences were presented to speakers in orthographic form in six separate lists, one for each fricative and preceding context (/t#f/, /ə#f/, /t#v/, /ə#v/, /t#z/, /ə#z/). Every list was organized into five blocks, with each block representing one repetition of the four

Table 1

Sentence sets for the sequences /t#f/, /t#v/, and /t#z/

Sentence /f/	Ich mag, was Anna gemalt hat. Felder und Wiesen sind auf dem Bild. I like what Anna drawn has. fields and meadows are on the picture 'I like what Anna has drawn. Fields and meadows are shown on the picture.'
Sentence /v/	Sie mag, was Benno gemalt hat. Wälder und Seen sind auf dem Bild. She likes what Benno drawn has. forests and lakes are on the picture 'She likes what Benno has drawn. Forests and lakes are shown on the picture.'
Sentence /z/	Er mag, was Clara gemalt hat. Senken und Hügel sind auf dem Bild. He likes what Clara drawn has. hollows and hills are on the picture 'He likes what Clara has drawn. Hollows and hills are shown on the picture.'
List /f/	Weil sie Gärten gemalt hat, Felder und Wiesen gemalt hat, und Bäume gemalt hat. Because she gardens drawn has, fields and meadows drawn has, and trees drawn has 'Because she has drawn gardens, fields and meadows , and trees.'
List /v/	Weil er Flüsse gemalt hat, Wälder und Seen gemalt hat, und das Meer gemalt hat. Because he rivers drawn has, forests and lakes drawn has, and the sea drawn has 'Because he has drawn rivers, forests and lakes , and the sea.'
List /z/	Weil sie Berge gemalt hat, Senken und Hügel gemalt hat, und Dünen gemalt hat. Because she mountains drawn has, hollows and hills drawn has, and dunes drawn has 'Because she has drawn mountains, hollows and hills , and dunes.'
Complement /f/	Weil sie vorhat, Felder und Wiesen zu malen, fährt sie aufs Land. Because she plans, fields and meadows to draw, goes she to-the countryside 'Since she wants to draw fields and meadows , she is going to the countryside.'
Complement /v/	Weil er vorhat, Wälder und Seen zu malen, fährt er nach Holstein. Because he plans forests and lakes to draw, goes he to Holstein 'Since he wants to draw forests and lakes , he is going to Holstein.'
Complement /z/	Weil sie vorhat, Senken und Hügel zu malen, fährt sie zum Aschberg. Because she plans hollows and hills to draw, goes she to-the Aschberg 'Since she wants to draw hollows and hills , she is going to the Aschberg.'
Word /f/	Anna hat Felder und Wiesen gemalt. Anna has fields and meadows drawn ' Anna has drawn fields and meadows.'
Word /v/	Benno hat Wälder und Seen gemalt. Benno has forests and lakes drawn ' Benno has drawn forests and lakes.'
Word /z/	Clara hat Senken und Hügel gemalt. Clara has hollows and hills drawn ' Clara has drawn hollows and hills.'

For the corresponding /ə/-context, *hat* 'has' was replaced by *hatte* 'had', and main verbs were replaced by their past tense form (e.g., Sentence /f/: Ich mochte, was Anna gemalt hatte. Felder und Wiesen waren auf dem Bild. 'I liked what Anna had drawn. Fields and meadows were shown on the picture.').

sentence types. The order of sentence types within blocks was pseudo-randomized and different for each block in a single list.

2.3. Recording procedure

Participants were familiarized with the test materials prior to the recording. They were asked to read the stimuli in a fluent and natural way, as if uttered as turns in informal conversation. Moreover, they were instructed to emphasize words printed in bold, but did not receive any specific instruction on prosodic phrasing.

Recordings were made in a sound-attenuated booth in the Phonetics department of Kiel University with a Sennheiser MD 421 microphone and the sound editing program CoolEdit. Signals were digitized directly into the computer at a sampling rate of 16 kHz. Participants read the test sentences from paper printouts at their normal speech rate. The experimenter monitored them throughout the recording, and asked for repetition of sentences in case of speech errors or accent-prosodic deviations from the intended realizations.

Table 2
Number of tokens by syntactic type and prosodic category

PCat:	MAJOR	MINOR	WORD	Total
Sentence	200	23	0	223
List	112	101	23	236
Complement	21	81	134	236
Word	0	9	230	239
Total	333	214	387	934

2.4. Prosodic labeling and prosodic categorization

The recordings were prosodically labeled by two phonetically trained native speakers of German. Pitch-accented target words (*Felder*, *Wälder*, *Senken*) were excluded from analysis. Preboundary pitch contours were categorized as continuant (no boundary tone/melodic break) on the one hand, or as rising (high boundary tone) or falling (low boundary tone) on the other hand. The categorization was based on auditory judgments and on f0 plots as provided by PRAAT (Boersma, 2001). It was carried out independently by each labeler, yielding an overall agreement of 98%. Ambiguous tokens were discussed with a third labeler. If no consensus was reached, the token was excluded from the data set.

We determined whether the target word was preceded by a pause. A pause was defined to be any portion of the speech signal before the onset of the target fricative that met one out of the following criteria:

- The fricative was preceded by any kind of filled pause (mostly breathing and incidentally prevoicing).
- In /ə/-context (/ə#_), there was silence between the offset of the /ə/ formant structure and frication onset, ranging from 22 to 947 ms, with a mean of 233 ms in our data set.
- In case of unreleased /t/ in /t/-context (/t#_), the /t/-closure before the following fricative was longer than 155 ms (i.e., more than four standard deviations from the mean), which is a portion of silence too long to be completely attributed to the /t/ (this was the case for 14 tokens, i.e., 1.5% of the data points, with the duration of silence ranging from 160 to 257 ms, with a mean of 195 ms).
- In /t/-context, the spectrogram suggested a segmentation of the release noise of /t/ followed by silence (26–885 ms, with a mean of 197 ms), and fricative onset.

Following the model of perceptual boundary strength by de Pijper and Sanderman (1994; see also Cho & McQueen, 2005), we assumed the presence of a pause by itself to be a sufficient cue to a major prosodic boundary. In our data, a pause was always accompanied by a boundary tone (BT), either low (L%) or high (H%). The prosodic category (PCat) MAJOR, roughly equivalent to an IP boundary, was thus defined as [+Pause, +BT]. Also in line with de Pijper and Sanderman, the absence of both a pause and a boundary tone ([-Pause, -BT]) was considered as a correlate of a weak boundary, equivalent to a prosodic word boundary (WORD). The intermediate boundaries, with a tonal break, but without pause, were classified as MINOR boundaries ([-Pause, +BT]). It should be noted that MINOR boundaries defined as such may include boundaries equivalent to IP and ip boundaries in a ToBI framework. Table 2 shows the token counts for prosodic categories by syntactic stimulus type for the analyzed data points (cf. Section 2.6 below).

2.5. Acoustic measurements

All acoustic measurements were based on simultaneous inspection of the waveform and the spectrogram of the speech signal as provided by PRAAT. Acoustic measurements included the duration of the preboundary syllable, which are /tə/ in *hatte*, and /hat/ in *hat*. In addition, we measured the duration of the target fricative

and the duration of voicing during the fricative.¹ If the preceding /t/ was released into the following fricative, /t/-aspiration and frication could not be separated and were both treated as part of the fricative. Since this was always the case at the two lower prosodic boundaries, pre- and postboundary segment duration measures differed for MAJOR versus MINOR/WORD prosodic categories in /t/-context. Thus, only MINOR versus WORD is directly comparable for this context. Voicing was defined as periodicity in the waveform, which was supported by the presence of a voice bar in the spectrogram.

In order to determine consistency, preboundary syllable duration, fricative duration and voicing during frication were measured a second time in a sample of 200 tokens, drawn from all speakers and conditions. The differences between the means of the original and the control measurements were 26 ms for the preboundary syllable duration (with a standard deviation of 32 ms), 3 ms for the fricative duration (with a standard deviation of 19 ms), and less than 1 ms for voicing (standard deviation of 12 ms).

2.6. Final data set

The data of two female speakers were entirely excluded from analysis, as they had not produced any MINOR prosodic boundaries. Moreover, 26 tokens produced by the remaining eight speakers were excluded for various reasons, such as pitch accents on target words, disagreement between the prosodic labelers, speech errors, consonant lenition of /v/ or /t/, or /t/-glottalization. The data set used for analysis thus consisted of 934 tokens.

3. Results

3.1. Domain-final lengthening

Preboundary lengthening is a well-established correlate of prosodic structure. We investigated whether our data were in accordance with previously reported evidence for final lengthening in German (e.g., Kohler, 1983; Wiese, 1996). Fig. 1 shows the average duration of the preboundary syllables /hat/ and /tə/ as a function of prosodic category and fricative.

Here and for all other analyses reported in this study, we built multi-level models with Speaker as random variable (Pinheiro & Bates, 2000; Baayen, 2004; Quené & van den Bergh, 2004). We opted for multi-level analyses because we expected and indeed found differences between speakers with respect to the number of prosodic domains they distinguish (cf. Fougeron, 2001). We modeled the duration of the preboundary syllable (/hat/, -/tə/) as a function of Prosodic Category (henceforth: PCat; MAJOR, MINOR, WORD), Fricative (/f, v, z/), and Context (/t, ə/).

All factors emerged as significant (PCat: $F(2,909) = 35.44$, $p < 0.001$; Context: $F(1,909) = 31.11$, $p < 0.001$; Fricative: $F(2,909) = 8.44$, $p < 0.001$). These main effects were modulated by three interactions: PCat interacted with Context ($F(2,909) = 10.60$, $p < 0.001$) and with Fricative ($F(4,909) = 2.92$, $p < 0.05$), and also the three-way interaction was significant ($F(4,909) = 3.33$, $p < 0.05$). Participants appeared to differ in their sensitivity to PCat and to Context, as indicated by the log-likelihood ratio's, which are measures indicating whether a more complex model outperforms a simpler model (Pinheiro & Bates, 2000). The log-likelihood ratio was 51.26, $p < 0.001$ for the comparison of the model with PCat as a random factor and the model in which participants were not allowed to differ in their sensitivity to any of the factors (i.e., where all factors were only treated as fixed factors). The log-likelihood ratio was 34.89, $p < 0.001$, for a model in which both PCat and Context were random variables, compared to a model where participants were allowed to only differ in sensitivity to PCat.

In order to investigate the interactions, we first split up the data for context. For both contexts, we found main effects of PCat and Fricative, and a significant interaction between these two factors (all $p < 0.05$). We

¹The RMS energy of the target fricative was also measured. However, we did not report the results for the following two reasons. First, the results did not show any effects as a function of prosodic boundary and assimilation context. Second, the RMS energy may not be the most reliable parameter for measuring the fricative strengthening, since the RMS energy or the intensity of the frication may also be influenced by voicing.

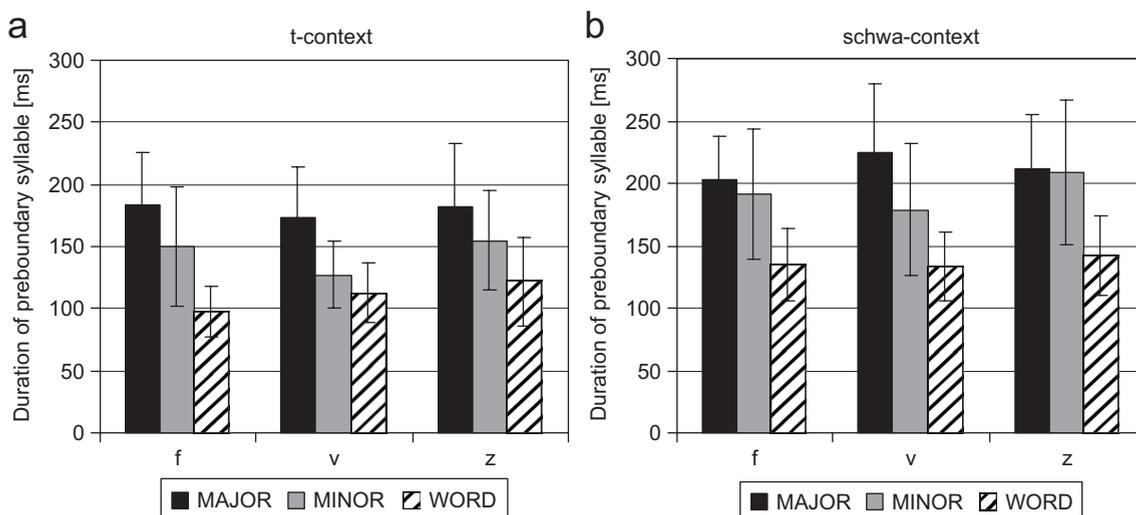


Fig. 1. (a) Duration of the preboundary syllable in /t/-context as a function of postboundary fricative (f,v,z) and prosodic boundary (MAJOR, MINOR, WORD). (b) Duration of the preboundary syllable in /ə/-context as a function of postboundary fricative (f,v,z) and prosodic boundary (MAJOR, MINOR, WORD).

then also split up the data by fricative. Individual comparisons showed that the final syllable /tə/ in ‘hatte’ was longer before a MAJOR and MINOR prosodic boundary than before a WORD boundary (all $p < 0.001$), without any difference between the MAJOR and MINOR boundary (all $p > 0.1$). The syllable /hat/ was the longest before a MAJOR boundary, and longer before a MINOR boundary than before a WORD boundary (all $p < 0.05$).

This analysis of preboundary syllable duration showed that there is a clear distinction between the MINOR and the WORD boundaries. The preboundary lengthening pattern supports the prosodic categorization based on the presence versus absence of a tonal break, the criterion by which we operationalized the distinction between the word level and the phrase level.

The distinction between the MAJOR and MINOR boundaries was not significant in /ə/-context, whereas it was in /t/-context. We have the impression that the difference attested in /t/-context has to be attributed to the segmentation procedure. Since the /t/-aspiration was inseparable from the fricative in cases where the two were not separated by a pause, the aspiration was regarded as part of the fricative in the MINOR and WORD conditions, but as part of the preboundary syllable in the MAJOR boundary condition (cf. Section 2.5 above). Thus, the inclusion versus exclusion of aspiration in the preboundary segment /t/ probably explains the difference between the MAJOR and MINOR boundaries.

In sum, our results confirm previous findings that preboundary lengthening is a phonetic correlate of an upcoming prosodic phrase boundary. We regard these results as evidence for the validity of our prosodic distinction between the MINOR and the WORD categories. Furthermore, it appears that preboundary lengthening by itself does not provide a criterion for distinguishing between major and minor IP in German.

3.2. Initial lengthening

We now turn to domain-initial lengthening and discuss the duration of the initial fricatives that cannot be affected by assimilatory devoicing, either because they are underlyingly voiceless, or because they occur in non-viable assimilatory devoicing context (i.e., are preceded by /ə/).

We first investigated the effect of PCat on the duration of the fortis fricative /f/. Since the /t/-release was attributed to the fricative duration except in the MAJOR condition (see Section 2.5 above), we only compared the MINOR category to the WORD category. We modeled the duration of the fricative as a function of PCat and Context. Only PCat appeared to be significant ($F(1,206) = 8.29, p < 0.05$). On average, /f/ was longer after

a MINOR phrase boundary than after a WORD boundary (135 vs. 115 ms). Speakers appeared to differ in their sensitivity to PCat (log-likelihood ratio = 26.56, $p < 0.001$): One participant did not show any effect of PCat at all.

Next, we analyzed the durations of the lenis fricatives /v, z/ in /ə/-context in all prosodic conditions. Fig. 2 illustrates the duration distributions of /v/ and /z/ in the three prosodic conditions by means of boxplots. The boxes show the interquartile range, the horizontal lines within the boxes indicate the median, and the vertical lines extend to observations within 1.5 times the interquartile ranges. Outliers beyond this range are plotted as individual circles. The analysis showed a main effect of PCat ($F(2,293) = 14.13$, $p < 0.001$) and of Fricative ($F(1,293) = 16.70$, $p < 0.001$), and an interaction of Fricative and PCat ($F(2,293) = 29.33$, $p < 0.001$).

In order to investigate the interaction, we ran separate analyses for the two fricatives. For both fricatives, we found an effect of PCat (/z/: $F(2,147) = 14.09$, $p < 0.001$; /v/: $F(2,139) = 19.99$, $p < 0.001$). Individual comparisons showed that for /z/, duration was longer (all $p < 0.01$) in the MAJOR and MINOR conditions (on average, 88 and 80 ms) than in the WORD condition (70 ms), but the difference between MAJOR and MINOR was not significant ($p > 0.05$). For /v/, fricative duration was the longest in MINOR (87 ms), followed by WORD (70 ms) and MAJOR (61 ms). All differences were significant (all $p < 0.01$). Unexpectedly, /v/-duration was shortest in the MAJOR condition. We come back to this finding in the General Discussion (Section 4).

Taken together, the duration analyses for /f/ and for non-assimilated /v/ and /z/ provide evidence for domain-initial strengthening in German. All segments following MINOR phrase boundaries are longer than word-initial, but phrase-medial segments.

3.3. Voicing

In this section, we discuss the effect of assimilation on the presence of vocal fold vibration, and the potential modulation by prosodic boundary size and phonotactic constraints.

We analyzed the percentage of voicing within each token, rather than absolute duration of vocal fold vibration, in order to normalize for fricative duration. The distribution of the percentages, as illustrated in Fig. 3, shows that 77% of the fricatives are produced either without (0%) or with complete (100%) vocal fold vibration. This is as expected, since /f/ is supposed to be voiceless, and indeed 81% of the /f/ tokens were produced without vocal fold vibration (the /f/-tokens thus forming 65% of all completely voiceless tokens). In addition, the lenis fricatives /v, z/ are expected to be produced with continuous voicing in intervocalic context without an intervening pause. Note that in this context, percentage scores close to, but smaller than 100% are physically improbable: Such percentage scores would imply an interruption of vocal fold vibration for just a

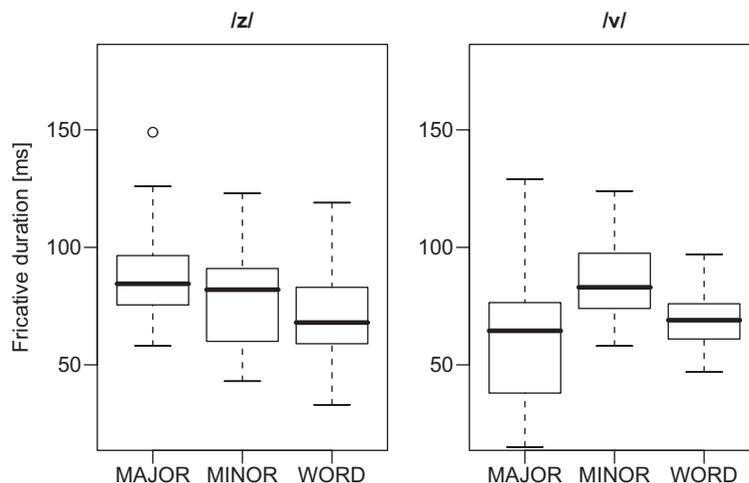


Fig. 2. Distribution of the durations of /z/ and /v/ in /ə/-context as a function of prosodic boundary (MAJOR, MINOR, WORD).

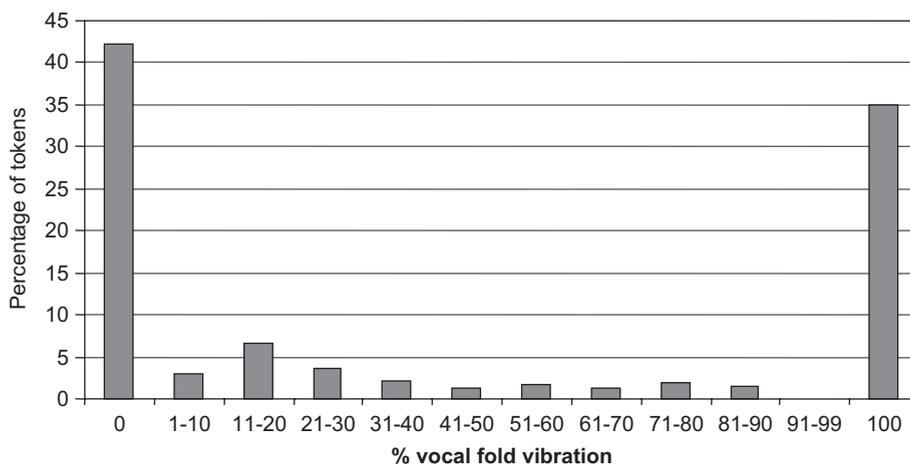


Fig. 3. Distribution of percentages of vocal fold vibration.

few milliseconds, which corresponds to a range from less than one, up to approximately three or four glottal cycles, depending on the speaker's f_0 and the given fricative duration. This accounts for why there are no tokens falling in the range of 91–99% in Fig. 3.

Since the percentage of glottal vibration does not show a normal distribution, we analyzed these data in two steps. We first investigated the likelihood of full voicing, that is, we compared the fully voiced fricatives with the not-fully voiced fricatives. Then we investigated the percentage of glottal vibration for fricatives that were not fully voiced.

3.3.1. Likelihood of full voicing

We first modeled the probability that a fricative was produced with full voicing. We excluded the /f/ data points from the analysis because we could not expect, and indeed did not find, any /f/ token produced with substantial vocal fold vibration. The predictors were Fricative (/v/, /z/), Context (/ə/, /t/), and Prosodic Category (MAJOR, MINOR, WORD). These predictors were entered in generalized linear multi-level models, assuming a binomial distribution of the data.

Both Fricative and Context emerged as significant main effects (Fricative: $F(1,602) = 9.35$, $p < 0.05$; Context: $F(1,602) = 31.39$, $p < 0.001$). The /z/ was realized with complete voicing in 45% of cases, and /v/ in 62%. As expected, most fully voiced fricatives occurred in /ə/-context (75% of all fricatives in /ə/-context, and 32% of all fricatives in /t/-context showed continuous glottal vibration).

There was an interaction of Context and PCat ($F(4,602) = 25.92$, $p < 0.001$). In order to investigate this interaction, we ran separate analyses for each context. For the /ə/-context, we found main effects of Fricative ($F(1,295) = 6.26$, $p < 0.05$) and of PCat ($F(2,295) = 42.82$, $p < 0.001$), which are illustrated in Fig. 4a. Separate analyses showed that all prosodic conditions differed significantly from each other (all $p < 0.05$). The larger the prosodic boundary, the less often the fricative was completely voiced.

We then turned to the /t/-context. Again, we found main effects for Fricative ($F(1,299) = 20.11$, $p < 0.001$), /v/ being more often fully voiced than /z/, and PCat ($F(2,299) = 9.17$, $p < 0.001$; see Fig. 4b). Separate analyses showed that there were more tokens of completely voiced fricatives after a MAJOR prosodic boundary (all $p < 0.001$) than after a MINOR or WORD boundary, which did not differ from each other ($p = 0.39$). The difference between MAJOR on the one hand and MINOR and WORD on the other hand is in line with the hypothesis that assimilatory devoicing is more likely to occur across lower prosodic boundaries.

3.3.2. Degree of assimilatory devoicing

Next, we investigated the fricatives that were produced with less than 100% voicing. We excluded /f/, as it showed only little glottal vibration. Furthermore, there were only a few tokens of /v/ and /z/ with less than 100% voicing in /ə/-context for the prosodic conditions MINOR and WORD, and therefore we analyzed the

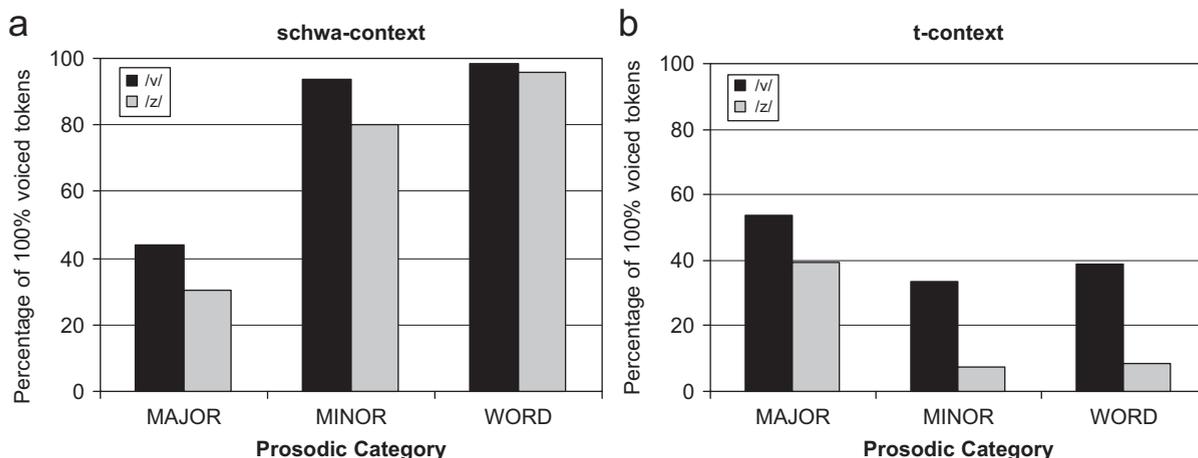


Fig. 4. (a, b) Percentages of /z/ and /v/ produced with 100% vocal fold vibration as a function of Prosodic boundary (MAJOR, MINOR, WORD) and context (/ə/, /t/).

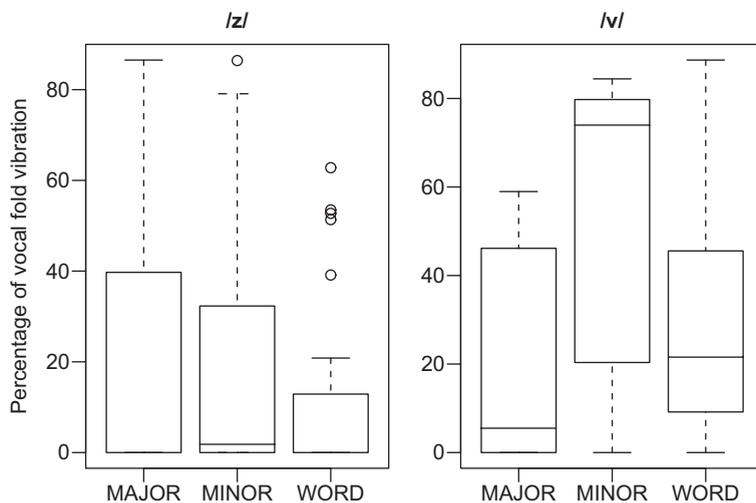


Fig. 5. Distribution of the percentages of vocal fold vibration for /z/ and /v/ in /t/-context as a function of prosodic boundary (MAJOR, MINOR, WORD).

/t/-context only. Fig. 5 shows the percentages of vocal fold vibration during /z/ and /v/ in /t/-context for the three prosodic conditions.

Again, we fitted a multi-level model with Speaker as random variable. This time, the percentage of voicing (calculated for each token separately, as the ratio of duration of vocal fold vibration by duration of the fricative) was the dependent variable. We found a significant effect of Fricative ($F(1,198) = 7.827, p < 0.05$) that differed slightly in effect size among participants (log-likelihood ratio = 6.82, $p < 0.05$). On average, tokens of /v/ were produced with a higher percentage of vocal fold vibration than tokens of /z/. Furthermore, PCat was significant ($F(2,198) = 7.09, p < 0.05$) and interacted with Fricative ($F(2,198) = 11.70, p < 0.001$).

We investigated the interaction by analyzing /v/ and /z/ separately. The analysis of /z/ showed a main effect of PCat ($F(2,117) = 6.91, p < 0.05$), and so did the analysis of /v/ ($F(2,74) = 12.24, p < 0.001$). For /z/, the difference between MAJOR and MINOR was not significant ($p > 0.05$), but these two differed significantly from WORD (all $p < 0.05$), which was the condition with the least amount of vocal fold vibration. For /v/, MINOR implied more voicing than MAJOR and WORD (all $p < 0.05$), and WORD showed a small, but significantly larger amount of voicing than MAJOR ($p < 0.05$).

3.3.3. Summary and discussion of the voicing data

Prosodic structure influenced voicing. In schwa-context, the likelihood of a lenis fricative to be completely voiced was larger after a Word boundary than after a Phrase boundary. Additional analyses were carried out to ascertain which acoustic properties correlating with PCat can account for this difference between the WORD and MINOR conditions. The likelihood of complete voicing was analyzed as a function of Fricative, PCat, Duration of the Fricative, and Duration of the preceding schwa. These analyses suggest that the duration of the schwa, rather than the duration of the fricative, overrides the effect of PCat. A longer schwa (i.e., more preboundary lengthening) appears to correlate with fewer completely voiced fricatives, suggesting that the fricative duration is not the only factor to explain the greater likelihood of devoicing.

For /t/-context, the percentage of vocal fold vibration during frication was larger in the MINOR condition than in the WORD condition (Fig. 5). These results support our hypothesis that assimilatory devoicing is stronger across lower prosodic boundaries.

As far as the MAJOR condition is concerned, the results suggest that the amount of glottal vibration is mainly determined by the difficulty to initiate phonation after a pause. For this condition, we found the least likelihood of complete voicing in /ə/-context and the lowest percentage of glottal vibration for /v/ in /t/-context. For the fricatives that were not completely voiced, vocal fold vibration was always absent at fricative onset in the MAJOR condition, independently of context. In the MINOR and WORD condition, in contrast, fricative onset was nearly always voiced in schwa context, but not in /t/-context.

Moreover, the results summarized in Fig. 5 suggest that in /t/-context, there was an equal amount of devoicing for /v/ and /z/ in the MAJOR condition. We therefore modeled the percentage of vocal fold vibration in all tokens of /v/ and /z/ (i.e., from both contexts, as there were enough tokens in this condition) which were produced with less-than-100%-voicing in the MAJOR condition, in order to test whether there were any effects of context or fricative, which were the predictors in the model. Both factors were far from significant (all $p > 0.1$). This indicates that the aerodynamic devoicing after a pause does not show the fricative effect, and occurs independently of the assimilation context.

Finally, we found a clear fricative effect associated with MINOR and WORD boundaries: /v/ is more likely to be fully voiced than /z/, and it devoices to a lesser degree. A phonetic explanation for these differences between /v/ and /z/ is that place of articulation has an impact on the aerodynamic conditions for voicing: Since the supraglottal cavity is larger for /v/ than for /z/, the transglottal pressure drop required for vocal fold vibration is easier to achieve and to maintain for /v/. In addition, it is equally possible that the functional load of the contrast of /v/ and /z/ with the corresponding fortis fricatives /f/ and /s/, and therefore of vocal fold vibration as a main cue to it, may play an important role. We come back to this issue in the General Discussion.

3.4. Interplay of initial lengthening and assimilation effects on duration

Three factors might influence the duration of /v/ and /z/ in /t/-context. First, there is initial lengthening, already shown for both /f/ and for /v/ and /z/ in /ə/-context (see Section 3.2). Such an initial lengthening effect may also be observed for /v/ and /z/ in /t/-context, yielding longer durations after phrase boundaries, in particular in the MINOR condition, than after word boundaries (the Initial Lengthening Hypothesis). Second, duration may be affected by assimilatory fortition, which may lengthen assimilated fricatives (the Fortition Lengthening Hypothesis). Since segments are likely to be more assimilated at a smaller boundary, one might predict that fricatives are produced with longer duration in the WORD than the MINOR condition, which is the opposite of the prediction of the first hypothesis. Third, duration may be affected by the overlap or reduction of articulatory gestures which may underlie assimilation. More gestural reduction and overlap would lead to shorter fricatives, resulting in shorter durations in the WORD than in the MINOR condition (the Overlap Shortening Hypothesis), as also predicted by the Initial Lengthening Hypothesis. In what follows we test these three hypotheses.

3.4.1. Lengthening or shortening in higher prosodic domains?

First, we investigated whether /v/ and /z/ in /t/-context are longer or shorter after larger prosodic boundaries. We modeled the duration of /v/ and /z/ in /t/-context as a function of PCat and Fricative. The

analysis showed a main effect of PCat ($F(2,297) = 7.27, p < 0.001$) and an interaction of Fricative with PCat ($F(2,297) = 20.31, p < 0.001$). Speakers differed in their sensitivity to Fricative (log-likelihood ratio = 16.49, $p < 0.001$, compared to a model in which participants did not differ in their sensitivity to any factor) and to PCat (log-likelihood ratio = 19.60, $p < 0.05$, compared to a model in which participants only differed in their sensitivity to Fricative). The random effect structure of the model suggested that only two participants showed the fricative effect.

We also analyzed the fricatives separately and found main effects of PCat on duration for both subsets ($/z/$: $F(2,149) = 10.80, p < 0.001$; $/v/$: $F(2,141) = 25.96, p < 0.001$). Fig. 6 shows the duration distributions for the two fricatives at the three prosodic boundaries. For $/z/$, there was no difference between the MAJOR (average: 80 ms) and the MINOR (83 ms) condition ($p > 0.1$), but each condition differed from WORD (66 ms; both $p < 0.001$) in favor of the Initial Lengthening and the Overlap Shortening Hypothesis. For $/v/$, the durational pattern was more complex: The longest duration was found for the MINOR condition (97 ms), and shortest duration for the MAJOR condition (59 ms), whereas the duration for WORD fell in between (69 ms, all $p < 0.001$). Note that these findings are similar to those found for the lenis fricatives in non-assimilatory $/ə/$ -context (Fig. 2), which suggests that assimilation did not affect the durations. Moreover, the shorter durations in WORD than in MINOR argue against a strong ‘fortition lengthening’ effect of assimilation.

3.4.2. Masked fortition lengthening?

The initial lengthening of the fricatives at higher prosodic boundaries may imply complete absence of fortition lengthening. However, one might argue that the fortition lengthening effect was unobservable in the previous analyses because it may have been masked by both initial lengthening and gestural reduction/overlap. We tested this possibility by investigating the correlation between fricative duration and percentage of voicing for $/v/$ and $/z/$. We hypothesized that, if there were fortition lengthening underlying the durational variation, segments would be longer if they were produced with less vocal fold vibration.

We restricted the data set to fricatives with a percentage of voicing larger than 0% and smaller than 100%, in order to obtain a normal distribution of this variable. We could not include both PCat and percentage of voicing as predictors for fricative duration, since PCat is an important predictor of percentage of voicing (see Section 3.3 above). We therefore ran separate analyses for the different prosodic conditions. Fig. 7(a)–(c) provides plots of fricative duration as a function of percentage of voicing. The solid lines represent non-parametric scatterplot smoothers (Cleveland, 1979). They suggest that there is a *positive* correlation in all three PCats, but only above 50% of voicing (i.e., the correlation appears to be non-linear), and we therefore included not only the percentage of voicing as a predictor into the model, but also its square. The model for

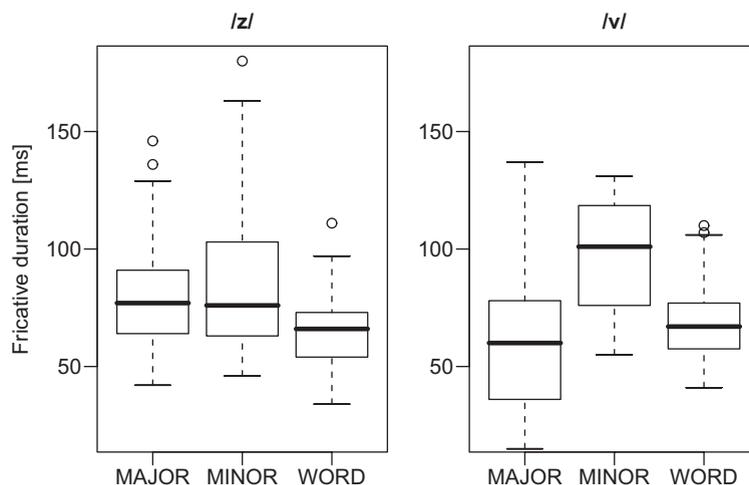


Fig. 6. Distribution of the duration of $/z/$ and $/v/$ tokens in $/ə/$ -context as a function of prosodic boundary (MAJOR, MINOR, WORD).

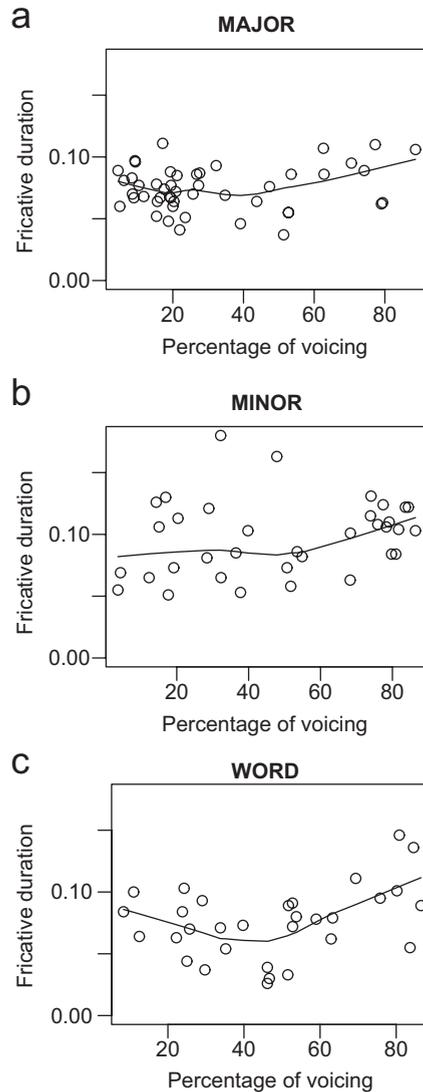


Fig. 7. (a–c) Duration of /v/ and /z/ as a function of percentage of voicing for the three prosodic boundaries (MAJOR, MINOR, WORD).

WORD showed no effect for percentage of voicing on fricative duration ($p > 0.1$). For MINOR, we attested only a marginally significant effect of the squared percentage of voicing ($F(1,26) = 3.01$, $p = 0.09$). Finally, for MAJOR, we found a significant effect for both percentage of voicing ($F(1,21) = 6.04$, $p < 0.05$) and its square ($F(1,21) = 6.01$, $p < 0.05$). In other words, in a condition where the strongest assimilatory devoicing force is present (i.e., in the WORD condition), we found no correlation between the fricative duration and the percentage of voicing; and where there is a correlation between the two variables, a higher percentage of voicing tends to be associated with longer duration, which is exactly the opposite of the prediction made by the Fortition Lengthening Hypothesis. The data thus suggests that fortition lengthening does not underlie the durational variation.

We offer the following explanation for the positive correlation observed in the MAJOR condition: Prosodic strengthening in /t/-context implies both longer duration (plain domain-initial strengthening) and less devoicing (resistance to assimilation). Also within a prosodic category, fricatives can differ in their strength. By consequence, some tokens are produced both with more vocal fold vibration and with longer duration than others.

3.4.3. Shortening as a result of gestural reduction and overlap?

The final question that we addressed with respect to the duration of the fricative was whether assimilated fricatives were shorter than non-assimilated ones (the Overlap Shortening Hypothesis). In a first analysis, we modeled the duration of /v/ and /z/ in the prosodic conditions MINOR and WORD, in both assimilatory (/t/-) and non-assimilatory (/ə/-) contexts. If assimilation shortened the fricatives, it would do so especially in the WORD condition in /t/-context, which would be reflected by an interaction of PCat with Context (i.e., a larger effect of PCat in /t/-context than in /ə/-context). However, this interaction was not significant ($p > 0.05$). In a second analysis, we investigated the duration of /v/ and /f/ in /t/-context in the same two prosodic conditions. Again if assimilation induced shortening, there would be an interaction of PCat with Fricative, as assimilatory shortening is not applicable to /f/. This interaction did not reach significance, either ($p > 0.05$). Our data therefore present no evidence for the Overlap Shortening Hypothesis.

4. General discussion

Studies on the role of prosodic structure on Voice Assimilation have been restricted so far to the effect of the presence versus absence of a word boundary and of neighboring words (e.g., Docherty, 1992; Slis, 1986). The present study is the first to investigate the effect of higher level prosodic structure on Voice Assimilation. The test case was Progressive Voice Assimilation in German. In a production study, we investigated prosodic effects on the duration of initial fricatives and on the relative duration of vocal fold vibration during the fricatives both in assimilation and non-assimilation context.

Fricatives were generally longer after higher prosodic boundaries, in both contexts. This finding of initial lengthening is in line with previous literature on domain-initial strengthening in other languages (Jun, 1998; Cho & Keating, 2001; Keating et al., 2003 for Korean; Cho & McQueen, 2005 for Dutch; Fougeron, 2001; Keating et al., 2003; Tabain, 2003a, 2003b for French; Fougeron & Keating, 1997 for American English; Keating et al., 2003 for Taiwanese). Thus, German can be added to the growing list of languages that show this subtle phonetic signature of prosodic structure at the left edge of prosodic domains.

The presence of a pause at MAJOR phrase boundaries blocked the initial lengthening process: Segments after a pause were found to be equally long (/z/), or to be even shorter (/v/) than segments after a phrase boundary without a pause (MINOR; Fig. 2). This finding suggests that speakers may choose among several options to indicate a high prosodic boundary. One option is to mark higher prosodic boundaries by domain-final and domain-initial lengthening, which can both be attributed to an underlying prosodic ‘ π -gesture’ slowing down the clock governing the articulatory gesture at a higher prosodic boundary (Byrd & Saltzman, 2003). Alternatively, speakers may combine preboundary lengthening and a pause, which is in itself a strong cue to a high prosodic boundary (de Pijper & Sanderman, 1994). Moreover, we found little vocal fold vibration in lenis fricatives after a pause, which may also function as a cue to a high prosodic boundary. Hence, in the presence of a pause, initial lengthening may not be a necessary additional cue.

Prosodic structure also affected the period of glottal vibration during the lenis fricatives /v/ and /z/. In non-assimilation context (/ə#_), the likelihood of partial devoicing was higher for fricatives initial in higher prosodic domains. Initiation of phonation may be delayed after the pause in the MAJOR condition. The difference between the MINOR and WORD condition is correlated with the duration of the preceding (preboundary) syllable, rather than with the duration of the fricative itself. We can think of no articulatory grounding for this observation.

A preceding voiceless obstruent (e.g., /t/) triggers assimilatory devoicing of /v, z/. We found that the degree of assimilatory devoicing depended on the size of the prosodic boundary between the domain-final obstruent and the domain-initial fricative: In /t/-context, the relative period of glottal vibration during the fricatives was shorter after lower prosodic boundaries, thus showing more assimilatory devoicing.

Effects of prosodic structure on sandhi processes have already been described within the framework of Prosodic Phonology (Nespor & Vogel, 1986). Sandhi processes are predicted to occur more frequently across lower than higher prosodic boundaries. In our data, however, the *likelihood* of devoicing in /t/-context was the same across phrase boundaries as across prosodic word boundaries (MINOR vs. WORD), but it was the *degree* of devoicing that differed between prosodic levels. Thus, our study suggests the extension of Prosodic Phonology, which currently focuses on the presence versus absence of phonological processes across

prosodic boundaries (categorical effects), to finer-grained subphonemic differences (gradient effects of prosodic structure).

Finally, we investigated the duration of fricatives in assimilation context (/t#_). As was the case in the non-assimilation context (ə#_), fricatives were longer after higher prosodic boundaries. Since fortis obstruents are typically longer and produced with less glottal vibration than lenis obstruents, it was expected that less glottal vibration due to assimilatory devoicing would be associated with longer duration (Fortition Lengthening Hypothesis). This appeared not to be the case. The data showed a positive, rather than a negative, correlation between the percentage of vocal fold vibration and fricative duration in the MAJOR condition and no correlation at all in the other conditions. Furthermore, we did not find any evidence for assimilation leading to shortening of the fricatives due to gestural reduction and overlap. In sum, the fricative durations were only determined by initial lengthening, and assimilatory devoicing did not affect them in either direction.

Data presented by studies on English voiceless plosives (Keating & Cho, 2005; Choi, 2003; Cole, Choi, Kim, & Hasegawa-Johnson, 2003) showed that both duration and voice onset time increase as the preceding prosodic boundary becomes larger. Hence, in English there are two cues to the feature [–voice] enhanced at higher boundaries. Interestingly, in the case of assimilatory devoicing of German fricatives, prosodic structure affects the two most important cues to the fortis-lenis distinction, duration and glottal vibration, in opposite directions. When lenis fricatives appear in prosodically stronger positions, their durations increase, which makes them more fortis-like. However, they also show a larger percentage of vocal fold vibration, which is a cue to lenis. In this sense, our data on German fricatives contrast with the findings for English voiceless plosives. We conclude that prosodic strengthening is not necessarily equal to ‘fortition’, as might be suggested by the findings concerning duration and VOT of English stop consonants, and by Fougeron’s (2001) argument that domain-initial consonants resisted lenition during sound change from Latin to French.

Since prosodic structure affects duration and glottal vibration in opposite directions, the implications of prosodic structure on the fortis-lenis distinction as perceived by the listener are still unclear. For computing the net effect of prosodic structure, we need to know the relative contributions of the two cues to perceived categorical membership. We can conclude, however, that the prosodic effect is smaller than it would have been if it had affected the two cues in the same direction. This suggests that the prosodic effects on the categorical identity of the fricatives are restricted in a way that results in a sufficient dispersion of fortis and lenis fricatives. The lenis fricatives always carry enough acoustic characteristics of lenis to be contrastive with their fortis counterparts, not only in voice assimilation context, but also after various prosodic boundaries.

Cho and McQueen (2005) observed a similar effect for Dutch plosives. They investigated /t/ and /d/ produced in three prosodic conditions comparable to those considered in the present study. Both /t/ and /d/ were longer after a MINOR boundary than after a WORD boundary. Since closure duration is typically longer for /t/ than for /d/ in Dutch, as in many other languages (e.g., Slis & Cohen, 1969a, b), boundary-induced durational variation might decrease the acoustic difference between the prosodically shortened versions of /t/ after a WORD boundary and the lengthened versions of /d/ after a phrase boundary. However, the phonemic contrast between /t/ and /d/ is always maintained at a given prosodic boundary. More importantly, the pattern of vocal fold vibration resulting from prosodic strengthening enhanced the language-specific implementation of the phonological contrast between /t/ and /d/ across prosodic levels: For /t/, vocal fold vibration restarted earlier after the burst following a larger than a smaller boundary (i.e., VOT was shorter after a larger prosodic boundary), while for /d/, prevoicing (i.e., negative VOT) was longer after a larger than a smaller boundary. After a larger boundary /d/ is more /t/-like with respect to its duration (initial lengthening), but the contrast with /t/ is enhanced by more prevoicing. Hence, the acoustic distance between /t/ and /d/ appears to be maintained, irrespective of variable strengthening due to prosodic structure. (See also Cho & Jun, 2000 for maintenance of the three-way stop contrast in Korean in different prosodic conditions).

We now turn to the differences between /v/ and /z/. Overall, /z/ was more devoiced than /v/. We offer two possible explanations for this difference. First, the intra-oral cavity is larger during the production of labiodental fricatives, which facilitates the maintenance of a constant transglottal pressure drop necessary for glottal vibration. This aerodynamic explanation is supported by the findings of Stevens et al. (1992), who investigated the devoicing of English /v/ and /z/ in several contexts and reported that 59% of the /v/-tokens, but only 38% of the /z/-tokens were voiced throughout.

The fricative effect may also be due to the lexical/phonotactic constraints of German, which has a functional contrast between /v/ and /f/, but not between /z/ and /s/ word-initially. If the difference between /v/ and /z/ is mainly driven by lexical/phonotactic constraints, we expect that /v/ is always substantially more lenis like than /z/ in one of the cues to the fortis lenis distinction, which is not necessarily the amount of glottal vibration. Important in this respect is the duration of /v/ in the MAJOR condition. We observed that /v/ was shorter in this condition compared to the MINOR condition, whereas we did not find a similar durational difference for /z/ (see Section 3.2, Fig. 2). A greater duration of /v/ in the MAJOR condition would have increased the competition with /f/, given that there is no compensation from a high proportion of vocal fold vibration to support the lenis identity. It is only if /v/ is more voiced, as in MINOR, that it can also be prosodically lengthened without becoming more confusable with /f/. In contrast, /z/, which does not need to compete with initial /s/, shows the durational increase in both the MINOR and MAJOR condition, irrespective of the amount of glottal vibration.

The influence of phonotactic constraints on assimilatory devoicing presupposes a cognitive component in assimilation processes (cf. Holst & Nolan, 1995; Nolan, Holst, & Kühnert, 1996). A follow-up question to be addressed is whether the fricative effect generalizes to all /v/-initial words, or is restricted to words which have an /f/-initial lexical competitor, as the stimuli in our experiment. In the latter case, assimilatory devoicing would be directed by the words actually stored in the speaker's mental lexicon (cf. Ernestus, Lahey, Verhees, & Baayen, 2006).

In summary, this study has shown that, in German, fricatives are longer at higher prosodic boundaries, irrespective of whether they are subject to assimilatory fortition. Assimilatory fortition only affects the percentage of vocal fold vibration in the fricative. Furthermore, phonotactic constraints allow lengthening effects to be stronger for the lenis fricative without a fortis counterpart. These findings demonstrate a complex interaction of prosody, sandhi processes, and phonotactic constraints. In spite of this interaction, the resulting speech signal nevertheless contains a set of clear acoustic cues to prosodic structure and to the fortis-lenis distinction.

Acknowledgements

This research has been supported by a grant from the Max Planck Gesellschaft zur Förderung der Wissenschaften. Thanks are due to Jonathan Harrington, Ernst Dombrowski and Herbert Fuchs (Kiel University) for providing facilities and technical support for the data collection, and to Marianne Radel for labeling the boundary tones. We thank Anne Cutler, James McQueen, Holger Mitterer, Michael Jessen and an anonymous reviewer for their helpful comments on previous versions of this paper.

References

- Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers 1*, 1, 1–45.
- Baer, T. (1975). Investigation of phonation using excised larynxes. Ph.D. dissertation, MIT.
- Beckman, M., & Pierrehumbert, J. (1986). *Intonational structure in Japanese and English*. *Phonology Yearbook 3*. (pp. 255–309). Cambridge: CUP.
- Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180.
- Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. Broe, & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 70–87). Cambridge: CUP.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149–180.
- Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 32, 141–176.
- Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America*, 117, 3867–3878.
- Cho, T., & Jun, S.-A. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society*, 36, 31–44 (An earlier version appeared in *UCLA Working Papers in Phonetics*, 99, 57–70).
- Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190.

- Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33, 121–157.
- Choi, H. (2003). Prosody-induced acoustic variation in English stop consonants. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 2662–2664). Barcelona, Spain.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829–836.
- Cole, J., Choi, H., Kim, H. & Hasegawa-Johnson, M. (2003). The effect of accent on the acoustic cues to stop voicing in radio news speech. In: *Proceedings of the 15th international congress of phonetic sciences* (pp. 2665–2998). Barcelona, Spain.
- Cole, R. A., & Cooper, W. E. (1975). Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, 58, 1280–1287.
- de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, 96(4), 2037–2047.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423–444.
- Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin: Foris.
- Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, 120, 1040–1051.
- Ferrero, F. E., Pelamatti, G. M., & Vaggies, K. (1979). Perceptual category shift of voiceless Italian fricatives as a function of duration shortening. In B. Lindblom, & S. Öhman (Eds.), *Frontiers of speech communication research. Festschrift for Gunnar Fant* (pp. 159–165). London etc.: Academic Press.
- Féry, C. (1993). *German intonation patterns*. Tübingen: Niemeyer.
- Fischer-Jørgensen, E. (1963). Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intraoralem Luftdruck. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 16, 19–36.
- Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetic*, 29, 109–135.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3728–3740.
- Fox, A. (1993). The prosodic hierarchy in German. In J. L. Flood, P. Salmer, O. Sayce, & C. Wells (Eds.), *Das unsichtbare Band der Sprache. Studies in German Language and Linguistic History in Memory of Leslie Seiffert* (pp. 455–481). Stuttgart: Heinz.
- Grabe, E. (1998). *Comparative intonational phonology: English and German*. (MPI Series in Psycholinguistics 7). Ph.D. dissertation, Nijmegen.
- Grice, M., & Baumann, S. (2000). Deutsche intonation und GToBI. *Linguistische Berichte*, 181, 1–33.
- Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics*, 6, 95–102.
- Holst, T., & Nolan, F. (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell, & A. Arvaniti (Eds.), *Papers in laboratory phonology IV* (pp. 315–333). Cambridge: Cambridge University Press.
- Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.
- Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, 85, 1718–1725.
- Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15, 189–226.
- Keating, P. A., & Cho, T. (2005). Influence of prosodic factors on segment articulations and acoustics in English. *Journal of the Acoustical Society of America*, 118, 2026.
- Keating, P. A., Cho, T., Fougeron, C., & Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 145–163). Cambridge: Cambridge University Press.
- Kissine, M., Van de Velde, H., & van Hout, R. (2003). An acoustic study of standard Dutch /v/, /f/, /z/ and /s/. *Linguistics in the Netherlands*, 93–104.
- Kohler, K. J. (1983). Prosodic boundary signals in German. *Phonetica*, 40, 89–134.
- Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 69–92). Dordrecht: Kluwer.
- Kohler, K. J. (1995). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.
- Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 7–30). Cambridge: Cambridge University Press.
- Mitterer, H. (2003). Understanding “garden bench”. Studies on the perception of assimilated word forms. Ph.D. dissertation, Maastricht.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- Nolan, F., Holst, T., & Kühnert, B. (1996). Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics*, 24, 113–137.
- Peperkamp, S. (1997). *Prosodic words*. (HIL Dissertations 34). Ph.D. dissertation, Amsterdam.
- Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II* (pp. 90–117). Cambridge: CUP.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS (statistics and computing)*. New York: Springer.
- Pirello, K., Blumstein, S. E., & Kurowski, K. (1997). The characteristics of voicing in syllable-initial fricatives in American English. *Journal of the Acoustical Society of America*, 101(6), 3754–3765.
- Quené, H., & van den Bergh, H. (2004). On multi-level modelling of data from repeated measures designs: A tutorial. *Speech Communication*, 43, 103–121.

- Selkirk, E. (1986). *On derived domains in sentence phonology*. *Phonology Yearbook 3* pp. 371–405. Cambridge: Cambridge University Press.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193–247.
- Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C. W., Price, P. J., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In , *Proceedings of the 1992 international conference on spoken language processing*, Vol. 2 (pp. 867–870).
- Slis, I. H. (1986). Assimilation of voice in Dutch as a function of stress, word boundaries, and sex of speaker and listener. *Journal of Phonetics*, 14, 311–326.
- Slis, I. H., & Cohen, A. (1969a). On the complex regulating the voiced-voiceless distinction I. *Language and Speech*, 12, 80–102.
- Slis, I. H., & Cohen, A. (1969b). On the complex regulating the voiced-voiceless distinction II. *Language and Speech*, 12, 137–155.
- Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America*, 91, 2979–3000.
- Tabain, M. (2003a). Effects of prosodic boundary on /aʃ/ sequences: acoustic results. *Journal of the Acoustical Society of America*, 113(1), 516–530.
- Tabain, M. (2003b). Effects of prosodic boundary on /aʃ/ sequences: articulatory results. *Journal of the Acoustical Society of America*, 113(5), 2834–2849.
- Uhmann, S. (1991). *Fokusphonologie: eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Tübingen: Niemeyer.
- van den Berg, R. J. H., & Slis, I. H. (1985). Perception of assimilation of voice as a function of segmental duration and linguistic context. *Phonetica*, 42, 25–38.
- Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, 22, 145–166.
- Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon Press.
- 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.
- Zsiga, E. C. (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, 28, 69–102.