

# Compensation for Assimilatory Devoicing and Prosodic Structure in German Fricative Perception

**Claudia Kuzla** <sup>\*a</sup>, **Mirjam Ernestus** <sup>b, a</sup> and **Holger Mitterer** <sup>a</sup>

<sup>a</sup> Max Planck Institute for Psycholinguistics Nijmegen

<sup>b</sup> Radboud University Nijmegen

**\*Address for correspondence:**

Claudia Kuzla

Max Planck Institute for Psycholinguistics

P.O. Box 310

6500 AH Nijmegen

The Netherlands

Email: [Claudia.Kuzla@mpi.nl](mailto:Claudia.Kuzla@mpi.nl)

Phone: +31 24 3521 334

Fax: +31 24 3521 213

## **Abstract**

An important source of phonetic variation in German fricatives is progressive voice assimilation: the lenis fricatives /v/ and /z/ are devoiced after /t/ across word boundaries. This process is gradient and moderated by prosodic structure: fricatives are more devoiced after smaller prosodic boundaries.

We present three phoneme identification experiments, investigating how listeners deal with assimilatory devoicing and its prosodic conditioning. Fully voiced, partially devoiced and completely devoiced fricatives had to be identified as fortis or lenis in different segmental (assimilation versus non-assimilation context) and prosodic (after a word versus a phrase boundary) environments. Results indicate that

1. listeners compensate for assimilatory devoicing in judging partially devoiced fricatives more often as lenis in assimilation context than in non-assimilation context;
2. prosodic structure plays a role in compensation for assimilation: more devoiced fricatives are more often judged as lenis after word boundaries than after phrase boundaries in assimilation context, and
3. the influence of prosody is constrained by lexical effects: we found prosodic conditioning of compensation for the devoicing of /v/, contrasting with /f/, but not of /z/. These findings suggest that an on-line prosodic analysis of spoken language contributes to the resolution of lexical ambiguity arising from progressive voice assimilation.

## 1. Introduction

Two major factors which may substantially alter the characteristics of speech sounds are sandhi processes and prosodic structure. These acoustic changes form a serious challenge for the listener, who (as it is traditionally assumed) has to map the altered speech sounds onto the word forms represented in the mental lexicon. Crucially, the two sources of variation are not independent from each other, since prosodic structure is known to constrain sandhi processes. This paper studies how listeners deal with the frequent sandhi process of assimilatory devoicing and its interaction with prosodic structure.

Assimilation processes have been well-studied in the linguistic literature. In phonology, assimilation is generally defined as the replacement of one speech sound by another. Recent research, however, has demonstrated the gradience of assimilation (Nolan 1992; Ernestus, Lahey, Verhees, and Baayen 2006; Snoeren, Hall↓, and Segui 2006; Kuzla, Cho, and Ernestus, in press). For instance, Nolan (1992) studied the assimilation of /t/ to a following /k/ in English. Electropalatographic recordings showed that the /t/ constriction may be realized as a complete alveolar closure (no assimilation), as a lateral closure (partial assimilation), or may be absent (completely assimilated to the velar constriction). A recent study by Kuzla et al. (in press) has shown the influence of prosodic structure on the degree of assimilation (see below).

Several factors facilitate the comprehension of assimilated segments. First, incomplete assimilation preserves acoustic cues to the original (unaltered) sound (e.g., Nolan 1992; Gow 2002; Mitterer 2003). Second, the comprehension of assimilated words depends on segmental context. Several studies have demonstrated, using a variety of experimental methods, that listeners easily achieve lexical access despite of assimilated segments, and do not even notice the difference between assimilated and

unassimilated sounds in contexts which license assimilation (e.g., Darcy 2002, to appear; Gow 2001, 2002; Gaskell 2003; Mitterer and Blomert 2003; Mitterer, Cs↓pe, and Blomert 2006). So far, there are no studies which investigated the effects of prosody on compensation for assimilation.

Prosodic structure as a source of pronunciation variation has been well established for many languages, including Dutch, English, French, German, Korean, and Taiwanese (Fougeron and Keating 1997; Jun 1998; Cho and Keating 2001; Fougeron 2001; Keating, Cho, Fougeron, and Hsu 2003; Cho and McQueen 2005; Kuzla et al. in press). In all these languages, segments at the beginning of prosodic domains are articulated more strongly than segments in medial position. This process is known as “domain-initial strengthening”. Importantly, domain-initial strengthening tends to be cumulative in the vertical dimension of the prosodic hierarchy. As a consequence, word-initial segments which are also phrase-initial are typically stronger than word-initial, but phrase-medial segments. Prosodic strengthening may affect phonetic characteristics of the speech sounds which also cue phonemic contrasts (e.g., duration, voice onset time, degree of constriction). Even if prosodic strengthening does not alter the phonemic identity of speech sounds, it may diminish the phonetic contrasts between variants of different phonemes across different prosodic positions.

So far, only one study has investigated the role of prosodic structure in speech perception above the word level (Cho, McQueen, and Cox 2007). In a series of cross-modal priming experiments, participants were auditorily presented with sentences containing two-word sequences, such as *bus tickets* (the primes), and performed lexical decision on the orthographic representation of the first word of these sequences (*bus*). The primes always contained a temporary ambiguity in that the first word plus the initial consonant of the second word (*bust*) formed an existing word, competing for recognition with the first word (*bus*). The two words of a prime were

realized such that they were separated by either a prosodic word boundary or an intonation phrase boundary. Crucially, the initial syllable of the second word (*ti* from *tickets*) was cross-spliced or identity-spliced over these two prosodic conditions. Listeners made faster lexical decisions on the first word if the initial syllable of the second word was cross-spliced from an intonation phrase onset rather than from a word onset within a phrase. The initial strengthening of the second word marked the lexical boundary between the two words more clearly and thus reduced the competition from words consisting of the target word followed by the initial consonant of the second word (e.g., *bust*). The authors also investigated whether initial strengthening had an effect on the recognition of the second word, but this appeared not to be the case.

In the present paper, we combine the research on the roles of assimilation and prosodic structure in speech perception. We studied listeners' identification of Standard German fricatives as fortis or lenis. The two lenis fricatives /v/ and /z/ may occur in word-initial position, as may the fortis fricative /f/. Fortis /s/, in contrast, is phonotactically illegal in word-initial position. Importantly, the two lenis fricatives are devoiced when preceded by voiceless obstruents. In such segmental contexts, the /z/ is claimed to be realized as [s] (Kohler 1990:79; 1995: 160). Thus *Sand* 'sand' is pronounced as [zant] in isolation, but is realized as [sant] in the sequence *hat Sand* 'has sand' ([hatsant]). This assimilation may be incomplete (Jessen 1998), that is, it may be gradient rather than categorical. /v/ is also assimilated after voiceless obstruents, but is claimed to remain distinguishable from /f/, on the one hand by being devoiced only partially in most cases, on the other hand by additional cues to lenis, such as less turbulence and shorter duration (Kohler 1995). Assimilation would thus never lead to confusion of /v/ and /f/, which function as different phonemes in word-initial position.

Kuzla et al. (in press) investigated how prosodic structure affects the realization of Standard German word-initial /f/, /v/ and /z/, and how prosodic strengthening interacts with the assimilatory devoicing of /v/ and /z/. The fricative-initial words *Felder* [fɛldɐ] ‘fields’, *Wälder* [vɛldɐ] ‘forests’, and *Senken* [zɛŋkən] ‘hollows’ were realized in full sentence utterances. They were preceded by different-sized prosodic boundaries, and occurred in assimilation and non-assimilation segmental contexts. For the present purpose, we only discuss the findings for the fricatives preceded by either a word boundary or a phrase boundary not accompanied by a pause. Prosodic structure appeared to affect the duration of the fricatives: they were longer after phrase boundaries than after word boundaries, in both assimilation and non-assimilation contexts. Assimilation did not affect the duration of the lenis fricatives, even though duration is an important cue to the fortis-lenis distinction. Both segmental context and prosodic structure affected the amount of glottal vibration during the lenis fricatives, as did the place of articulation of the fricative. As expected, fricatives in assimilation context were realized with less glottal vibration than fricatives in non-assimilation context. More importantly, the degree of assimilatory devoicing was larger after prosodic word boundaries than after phrase boundaries. Finally, /z/ was devoiced to a greater extent than /v/. Kuzla et al. (in press) explained the difference between the two fricatives partially on aerodynamic and partially on functional grounds. Maintenance of glottal vibration may be easier for /v/ than for /z/, because of the larger oral cavity. Additionally, in Standard German the phonemes /v/ and /z/ differ in their functional load: While the contrast between /v/ and its voiceless counterpart /f/ encodes lexical distinctions in word onset position, the /z/-/s/ contrast does not, as /s/ is phonotactically illegal in word-initial position. As a consequence, speakers may only fully devoice word-initial /z/, because complete devoicing of /v/ might lead to stronger lexical competition with /f/-initial words. As this production

study only investigated the realization of a /v/-initial word forming a minimal pair with an /f/-initial word, it is unclear whether this effect of place of articulation is due to this existing /f/-initial word (lexical effect), or just to the possibility that there may be /f/-initial words, but no /s/-initial words (phonotactic effect). Note, however, that a phonotactic constraint always implies a lexical constraint, but not vice versa. In the following, we will not distinguish between the two constraints and refer to both of them as lexical constraints.

It is important to note that the assimilation process under investigation is rather unusual in affecting a phoneme in onset position (cf. Beckmann, 1997). Models of word recognition rely heavily on correct onset perception for creating shortlists (Norris 1994) or cohorts (Gaskell and Marslen-Wilson, 1997). This underscores the importance of compensation for progressive assimilation.

We report here three categorization experiments investigating how listeners compensate for the effects of prosodic structure and assimilation on the realization of Standard German fricatives. We presented listeners with sentences containing target words of which the initial fricatives had been manipulated. Listeners were asked to identify the target word as starting with a lenis or fortis fricative (2-Alternative Forced Choice), which is an offline task. Importantly, 2AFC results for compensation for assimilation have been replicated with online measures (e.g., Mitterer and Blomert 2003).

Since Kuzla et al. (in press) have shown a complex interaction of assimilation, prosodic structure, and lexical constraints especially for the amount of glottal vibration, the crucial manipulation was the amount of glottal vibration during the fricative.

In the three experiments, we tested the following hypotheses. The first hypothesis concerns the role of segmental context. The production study has shown

that assimilation context substantially reduces the amount of glottal vibration. Since glottal vibration is an important cue to the fortis-lenis distinction, assimilated fricatives are thus more fortis-like. Hence, recognition of a lenis fricative may be seriously hindered in assimilation context, unless listeners compensate for assimilatory devoicing. In line with previous findings, we hypothesize that listeners compensate for assimilatory devoicing, as they also do for place assimilation (e.g., Gow 2001, Gow and Im 2004; Coenen, Zwitserlood, and Bólte 2001, Darcy 2002, to appear, Snoeren et al. 2006). This implies that less glottal vibration will be necessary for a lenis percept in assimilation context than in non-assimilation context.

Our second hypothesis concerns the role of prosodic structure. It appeared from Kuzla et al. (in press) that prosodic structure modulates the effect of assimilation. Fricatives were more assimilated, that is, more devoiced, after a word boundary than after a phrase boundary. If listeners compensate for assimilation, we may therefore expect that they compensate for more devoicing after a word boundary than after a phrase boundary. In other words, we hypothesize that a fricative requires less glottal vibration for a lenis percept after a word boundary than after a phrase boundary.

The first two experiments focus on the contrast between /v/ and /f/, which differentiates between words. Experiment 3 is concerned with the /z-s/ contrast, which is lexically irrelevant in initial position. In production, the difference in functional relevance between the two fortis-lenis contrasts and the difference in the aerodynamic configurations for the two places of articulation result in more assimilatory devoicing for /z/ than for /v/. Given that /z/ is produced with more variation, we may formulate two opposite hypotheses on the compensation for assimilation. On the one hand, listeners may have formed expectations about the degree of glottal vibration in alveolar fricatives. This may lead listeners to compensate more for devoicing of /z/



than of /v/, since /z/ is often produced with less glottal vibration. On the other hand, listeners may completely ignore the amount of glottal vibration for /z/ and thus show no compensation for assimilation at all, since they do not need to distinguish between /z/ and /s/ in onset position.

## **2. Experiment 1: Perception of devoiced /v/**

In Experiment 1, we investigated how the amount of vocal fold vibration during the labiodental fricative /v/ codetermines the perception of this fricative as lenis or fortis in four conditions: in assimilation and non-assimilation context, after a prosodic word boundary and a phrase boundary. Importantly, the duration of the fricative did not covary with the amount of vocal fold vibration, but was adjusted to the preceding prosodic boundaries: the fricatives were longer after phrase boundaries than after word boundaries, reflecting domain-initial strengthening.

### **2.1. Participants**

Sixteen native listeners were paid to participate in the experiment. They were all students at Kiel University. None of them reported any hearing problems.

### **2.2. Stimuli**

We created four sentences containing the fricative-initial target word *Wälder* [vɛldɐ] 'forests', which forms a minimal pair with *Felder* [fɛldɐ] 'fields' (see Table 1). The sentences were semantically unbiased towards *Wälder* or *Felder*. The word-initial fricative was preceded by either *hat* 'has', ending in the assimilation context /t/, or by *hatte* 'had', ending in the non-assimilation context /ə/. These context + target sequences occurred in two different syntactic constructions which induced different prosodic boundaries between the context word and the target word: a prosodic phrase

boundary, marked by a boundary tone and preboundary lengthening, or a prosodic word boundary.

Table 1 about here

We recorded a female native speaker reading several repetitions of the four sentences and a number of similar sentences. The recordings were made in a sound-attenuated booth, by means of a Sennheiser microphone, and directly stored onto a computer, at a sample rate of 48 kHz. We instructed the speaker to realize a contrastive accent on another word in the sentence (indicated in bold in Table 1) than the target word, in order to avoid accentuation of this target word.

From these recordings, we generated two fricative continua ranging from fully voiced /v/ to completely devoiced /v/, using PRAAT ([www.praat.org](http://www.praat.org)). In one continuum, the duration of the fricative was appropriate for a word boundary, while in the other continuum, this duration was adjusted to a phrase boundary (see below).

For the word boundary continuum, we selected a recording of the sentence *Anna hatte Walder und Wiesen gemalt* where the initial /v/ in *Walder* was produced with continuous glottal vibration ( $f_0 = 150$  Hz), a mean RMS amplitude of 0.007 Pa, and a duration of 50 ms, which appeared to be prototypical for this speaker. This sound was used as the fully voiced endpoint of the continuum. From a recording of the sentence *Anna hat Walder und Wiesen gemalt* (the assimilation context), we selected a completely devoiced realization of /v/, which served as the voiceless endpoint of the continuum. This sound was equalized in duration with the fully voiced endpoint by cutting 0.7 ms of frication from the middle of the fricative. We created intermediate steps of the continuum by replacing the glottal cycles of the continuum's voiced endpoint one by one by voiceless frication from the continuum's voiceless

endpoint, starting from the left (cf. Figure 1). Given that the voiced endpoint contained 8 glottal cycles, this resulted in 8 continuum steps in total.

Figure 1 about here

For the phrase boundary continuum, we selected a completely voiced /v/ from a token of *Wälder* in the utterance *Weil sie vorhatte, Wälder und Wiesen zu malen, fuhr sie aufs Land*, which had a duration of 77 ms. RMS analyses indicated that the amplitude of this sound (0.008 Pa) was almost equal to the voiced endpoint of the Word boundary continuum (0.007 Pa).

Since we had not elicited any completely devoiced /v/ in the phrase condition, we lengthened the completely devoiced /v/ from the word condition using PSOLA, until it also had a duration of 77 ms. Hence, the two unvoiced endpoints were exactly equal in amplitude for the two prosodic conditions (mean RMS: 0.001 Pa). Then we replaced the glottal cycles of the completely voiced realization one by one with voiceless frication from the completely voiceless token, starting from the left. Whereas the fundamental frequency was constant in the fully voiced /v/ in the word condition, it fell from 288 Hz to 230 Hz during the /v/ in the phrase condition, reflecting the pitch fall from the high boundary tone aligned with the preceding syllable. Consequently, glottal cycles became longer from the left end to the right end in the completely voiced sound, and therefore also in the intermediate steps. Hence, we obtained shorter step widths towards the voiced end of the phrase continuum. This explains the disproportionately larger number of 18 continuum steps (versus 8 for the Word continuum), which are not explained by the 50% increase in duration (77 ms versus 50 ms).

In order to generate comparable continua, we calculated the percentage of

glottal vibration for each step within each prosodic condition, and selected 7 steps from each continuum, such that corresponding steps of the two continua were matched as closely as possible in the percentage of glottal vibration during the fricative. This percentage ranged from 0% to 84% in equidistant steps of 14% ( $\forall 1\%$ ). The 100% voiced endpoints were excluded, so that the continua contained neither prototypical /v/ nor /f/ sounds. Having a good, 100% voiced /v/, but no good /f/ in the stimuli might have given rise to a range effect (cf. Repp & Liberman, 1987).

The test sounds were spliced into the four sentences which contained the target word *Wlder* and represented the four experimental conditions that result from crossing the factors Context (assimilation context, non-assimilation context) and Prosodic Boundary (Phrase, Word) (see Table 1). This yielded (4 conditions \* 7 continuum steps =) 28 experimental stimuli.

### **2.3. Procedure**

Participants were seated in a quiet seminar room in front of a portable computer. They were presented via headphones with each auditory test stimulus ten times, in a randomized order which was different for each participant. Simultaneously, the participants saw the words *Felder* and *Wlder* on the screen.

Participants were instructed that they would hear sentences containing manipulated versions of the words *Felder* and *Wlder*, and that they would probably not encounter “perfect” instances of either word. They were asked to indicate as fast as possible whether a sentence contained *Felder* or *Wlder* by pressing the response button labeled ‘F’ for *Felder* and the response button labeled ‘W’ for *Wlder*. We did not analyze the reaction time, here and in all following experiments, as the mean RT, measured from fricative onset, always exceeded 900 ms, indicating the offline nature of the task.

After four training trials, the experiment started. After every block of 50 trials, participants were allowed to take a short break. An experimental session lasted about 20 minutes.

## 2.4. Results and Discussion

Figure 2 shows the percentages of /v/-identifications averaged over Prosodic Boundary, Context and Step. We also computed these averages for each participant separately and analyzed these ( $2 \times 2 \times 7 \times 16 =$ ) 448 averages by means of a 3-way repeated measures ANOVA, with as independent variables Boundary (Word versus Phrase), Context (assimilation versus non-assimilation), and Step (1-7). All three factors emerged as significant (Boundary:  $F(1, 15) = 6.482$ ,  $p < 0.05$ ; Context:  $F(1, 15) = 10.887$ ,  $p < 0.01$ ; Step:  $F(6, 90) = 7.030$ ,  $p < 0.001$ ). In addition, the interaction of Boundary by Context ( $F(1, 15) = 11.145$ ,  $p < 0.01$ ) and of Context by Step ( $F(6, 90) = 9.347$ ,  $p < 0.001$ ) were significant, whereas the three-way interaction was not ( $F(6, 90) = 0.682$ ,  $p = 0.67$ ).

Figure 2 about here

In order to investigate the interaction of Boundary and Context in more detail, we first analyzed the two contexts separately. For the non-assimilation context, we found main effects of Boundary ( $F(1, 15) = 11.504$ ,  $p < 0.01$ ) and Step ( $F(6, 90) = 15.656$ ,  $p < 0.001$ ), but no interaction. The percentage of /v/-responses increased towards the voiced end of the continuum (from 52% to 79%, averaged over the two boundary conditions; see Figure 2). More surprisingly, more /v/-responses were given in the Word condition than in the Phrase condition (75% versus 57%, averaged over steps). For the assimilation context, neither Boundary nor Step was significant, nor

was the interaction.

We also analyzed the two boundary conditions separately to see whether there was an effect of context. In the Phrase condition, there were main effects of Context ( $F(1, 15) = 13.531, p < 0.01$ ), which implied more /v/-responses in assimilation context than in non-assimilation context (81% versus 57%, averaged over steps), and of Step ( $F(6, 90) = 4.211, p < 0.01$ ), yielding more /v/-responses towards the voiced end of the continuum (from 60% to 76%, averaged over contexts). The interaction of Context by Step was also significant ( $F(6, 90) = 4.546, p < 0.001$ ). Individual pairwise comparisons of the two contexts for all steps showed that the context effect was present mainly in the more devoiced part of the continuum (steps 1, 2, 3, and 6; all  $p < 0.05$ ). In the Word condition, we also found main effects of Context (Assimilation context: 88% versus non-assimilation context: 75%, averaged over steps;  $F(1, 15) = 6.398, p < 0.05$ ) and Step (from 76% to 86%, averaged over contexts;  $F(6, 90) = 2.771, p < 0.05$ ), and a significant interaction ( $F(6, 90) = 4.559, p < 0.001$ ). Pairwise comparisons revealed that the Context effect was again restricted to the devoiced end of the continuum, and was absent for the steps 5, 6, and 7.

These results present convincing evidence for compensation for assimilation. Listeners judge the same acoustic signal more often as /v/ in assimilation context than in non-assimilation context. We did not find evidence for a role of prosodic structure in compensation for assimilation: Listeners identified all fricatives in assimilation context equally often as /v/, irrespective of prosodic boundary. Possibly, this reflects a ceiling effect: due to the compensation for assimilation driven by the /t/-context, nearly all stimuli were already perceived as /v/.

The effect of prosodic structure observed for the non-assimilation context is somewhat surprising. In production, partial devoicing in this context is more common in the Phrase condition than in the Word condition (Kuzla et al. in press). If there

were perceptual compensation for this type of devoicing, we would expect more /v/-responses in the Phrase condition than in the Word condition. However, we found the opposite. Since duration is also an important cue to the fortis-lenis distinction, with fortis fricatives being longer, a possible explanation for the present result may be that the fricatives were much shorter in the Word condition than in the Phrase condition, leading to more /v/-responses in the Word condition.

### **3. Experiment 2: Perception of the /v/ - /f/ contrast**

In Experiment 2, we investigated again whether prosodic structure moderates compensation for assimilatory devoicing. The durations of the fricatives within the two continua in Experiment 1 were the same for all steps of the voicing continuum within one prosodic domain. These durations were those of the lenis fricatives and overridden a prosodic effect on the compensation for assimilatory devoicing. In Experiment 2, we covaried the durations of the fricatives with their amount of vocal fold vibration, such that the test sounds reflected the natural fortis-lenis contrast better.

#### **3.1. Participants**

Twenty native listeners, all students at Kiel University, who had not participated in Experiment 1, were paid to participate in Experiment 2. None of them reported any hearing disorders.

#### **3.2. Stimuli**

A second female native speaker of Standard German recorded the same materials as

the first speaker in Experiment 1, plus the sentences given in Table 2 with the target word *Felder* ‘fields.’ The recording procedure was the same as in Experiment 1.

Table 2 about here

We generated one /v-f/ continuum from natural endpoints. As the /f/-endpoint, we selected a completely voiceless token of /f/ from a *Felder* sentence with a Phrase boundary before the fricative-initial target word in non-assimilation context (*Weil sie vorhatte, Felder und Wiesen zu malen, fuhr sie aufs Land*). This sound had a duration of 122 ms and an RMS of 0.002 Pa. The /v/-endpoint was a completely voiced token of /v/ from a phrase-initial realization of *Wälder* in non-assimilation context, which contained 20 glottal cycles, had an RMS amplitude of 0.004 Pa, and a duration of 68 ms. Note that despite the different speaker, this voiced endpoint is very similar in fundamental frequency, RMS, and duration to the fully voiced continuum endpoint in the Phrase condition in Experiment 1. We then created 18 intermediate steps by replacing the glottal cycles of the /v/-endpoint one by one by a part of the /f/-endpoint, starting from the left. Since the /f/-endpoint was 54 ms longer than the /v/-endpoint, the left (/f-) part of every intermediate step was  $(54/20 = )$  2.7 ms longer than the voiced signal portion which it replaced (see Figure 3).

Figure 3 about here

We created two sets of stimuli. In the first set of stimuli, the 20 steps replaced the initial /v/ of four tokens of *Wälder*, realized in the four context-plus-boundary conditions (Table 1). Our speaker always produced some anticipatory voicing in the /t/ preceding *Wälder*, which may bias perception towards /v/ and thus mask



compensation for prosodic structure. In the second set, the 20 steps of the continuum replaced the initial /f/ of two tokens of *Felder*, one realized after a /t/ in phrase-initial position, the other one realized after a /t/ in word-initial position (see Table 2). The absence of anticipatory voicing in these sentences is a cue neither to /v/ nor /f/, as anticipatory voicing is often absent before /v/ for many speakers. Hence, sentences containing a completely voiceless /t/, as produced by our speaker before *Felder*, may reveal a stronger effect of prosodic structure on compensation for assimilation.

### 3.3. Procedure

The procedure was the same as in Experiment 1, except that, due to the larger number of experimental stimuli, each stimulus was presented only four (instead of ten) times.

### 3.4. Results

We first analyzed the data from the first set of stimuli. Figure 4 shows the percentages of /v/-responses for the 20 steps in the four context plus prosodic boundary conditions.

Figure 4 about here

As in Experiment 1, we averaged the /v/-responses for each participant over context (/t/ versus /ə/), prosodic boundary (Word versus Phrase) and continuum step (1 through 20). We analyzed these averages by means of a 3-way repeated measure ANOVA. All three factors emerged as significant (Boundary:  $F(1, 19) = 39.30$ ,  $p < 0.001$ ; Context:  $F(1, 19) = 70.449$ ,  $p < 0.001$ ; Step:  $F(19, 361) = 267.658$ ,  $p < 0.001$ ). In addition, the two-way interactions of Boundary by Context ( $F(1, 19) = 53.225$ ,  $p < 0.001$ ), Boundary by Step ( $F(19, 361) = 6.366$ ,  $p < 0.001$ ) and of Context by Step

( $F(19, 361) = 11.916, p < 0.001$ ) as well as the three-way interaction of Boundary by Context by Step ( $F(19, 361) = 4.486, p < 0.001$ ) were significant.

In order to investigate the 3-way interaction, we split the data by Boundary. First, we analyzed the Phrase condition. There were significant main effects of Context ( $F(1, 19) = 41.962, p < 0.001$ ) and of Step ( $F(19, 361) = 174.216, p < 0.001$ ), and also the interaction was significant ( $F(19, 361) = 4.623, p < 0.001$ ). Pairwise comparisons of the two contexts for each step showed that there were significantly more /v/-responses in the assimilation context (/t/) than in the non-assimilation context (/ə/) for the steps 6 -13, that is, for the middle part of the continuum.

Also in the Word condition, the main effects of Context ( $F(1, 19) = 77.895, p < 0.001$ ) and of Step ( $F(19, 361) = 155.809, p < 0.001$ ) were significant, in addition to the interaction ( $F(19, 161) = 11.550, p < 0.001$ ). Pairwise comparisons for each step revealed that the Context effect (more /v/-responses in assimilation than in non-assimilation context) was significant for the first 15 steps, representing the stimuli with maximally 75% of glottal vibration. The compensation for assimilation is thus extended to the voiceless part of the continuum in the Word condition compared to the Phrase condition. This supports the hypothesized effect of prosodic structure on compensation for assimilation. Assimilation may devoice a larger part of the fricative in the Word condition than in the Phrase condition. As a consequence, listeners accept (almost) completely devoiced fricatives more readily as realizations of /v/ after Word boundaries than after Phrase boundaries.

The finding that the completely voiceless stimulus (step 1) was in almost 20% of the cases interpreted as /v/ in the Assimilation plus Word condition (see Figure 4) suggests that listeners interpreted the anticipatory voicing in the preceding /t/ and potentially also the characteristics of the following vowel as cues to the original voicing of the fricative (/v/).

Figure 5 about here

We therefore also analyzed the responses to the second set of stimuli, where this anticipatory voicing was absent, and any cue in the vowel would be in line with a fortis interpretation. Recall that the stimuli in the second set consisted of the same /f-v/ continuum, but the test sounds replaced the initial fricative of *Felder*, instead of *Walder*, produced in /t/-context in the two prosodic conditions. Figure 5 shows the percentages of /v/-judgments for the 20 steps in these two conditions. Different from Figure 4, the categorization function now approximates 0% lenis (=100% fortis) judgments towards the /f/ endpoint. This shows that the slightly voiced /t/ and possible other cues in the following vowel affected the fortis/lenis judgments in the first stimulus set. In the second set, we still found a strong effect of prosody in a two-way repeated measure ANOVA with Prosodic Boundary and Step as independent variables (note that the context was always /t/ in this data set). Again, Boundary ( $F(1, 19) = 104.610$ ,  $p < 0.001$ ) and Step ( $F(19, 361) = 159.510$ ,  $p < 0.001$ ) emerged as significant, as did the interaction ( $F(19, 361) = 16.530$ ,  $p < 0.001$ ). Analyses of the separate steps showed that significantly more /v/-responses were given in the Word condition than in the Phrase condition for the steps 4-17. As for data set 1, less glottal vibration was required for a lenis percept after a Word boundary than after a Phrase boundary. Hence, this data set also shows an effect of prosodic structure on compensation for assimilation.

Figure 5 about here

In conclusion, Experiment 2 shows compensation for assimilation, and the

hypothesized effect of prosodic structure on this compensation. We have also found compensation for assimilation in Experiment 1, but that experiment did not reveal an effect of prosodic structure. In Experiment 2, we made use of a complete /f-v/ continuum by covarying the duration of the fricative with the amount of glottal vibration. By doing so, we did not obtain ceiling effects as we probably did in Experiment 1.

#### **4. Experiment 3: Perception of the /z/ - /s/ contrast**

Experiments 1 and 2 focused on the /v-f/ contrast, which is lexically distinctive in word-initial position in Standard German. In Experiment 3, we investigated the /z-s/ contrast, which distinguishes words only in word-medial position in this language. In the introduction, we formulated two hypotheses regarding a potential influence of lexical constraints on compensation for assimilation. On the one hand, listeners may have formed expectations about the degree of glottal vibration in alveolar fricatives. They may expect /z/ to be devoiced more strongly than /v/ (Kuzla et al. in press), and therefore compensate more for devoicing of /z/ than of /v/. On the other hand, listeners may completely ignore the amount of glottal vibration in /z/, as they do not need to distinguish between the fortis /s/ and lenis /z/.

##### **4.1. Participants**

The participants were the same as those for Experiment 2. The order of experiments was counterbalanced across participants, and there were at least two days between the two experimental sessions for each participant.

## 4. 2. Stimuli

We used materials similar to those in the previous experiments. The sentences for this experiment (Table 3) contained either the target word *Senken* [zɛŋkən] ‘hollows’ or the non-word [sɛŋkən]. Given that alveolar fricatives in onset position are voiced, if not preceded by a segment licensing assimilatory devoicing, we can assume that [z] is the naturally produced allophone in ‘/z/enken’ in non-assimilation context. To obtain unvoiced tokens, we instructed the speaker explicitly to produce an [s] by exemplifying the /s-z/ contrast in word-medial position, which occurs in Standard German.

Segmental context and prosodic environment were again varied resulting in the same four conditions as in experiments 1 and 2: 1. Assimilation context /t/ plus Word boundary, 2. Assimilation context /t/ plus Phrase boundary, 3. Non-assimilation context /θ/ plus Word boundary, and 4. Non-assimilation context /θ/ plus Phrase boundary. The materials were recorded in the same session as the materials for Experiment 2 and by the same female speaker.

Table 3 about here

We selected a fully voiced token of /z/, which had a duration of 74 ms, an RMS amplitude of 0.008 Pa, and contained 20 glottal cycles, in addition to a completely voiceless /s/, which had an RMS of 0.015 Pa and a duration of 145 ms. These sounds came from phrase-initial realizations of the word /z/enken and the non-word \*/s/enken, respectively, both produced in non-assimilation context. We created from these two endpoints a 20-step continuum as in Experiment 2. The 20 test sounds then replaced the initial /z/ of /z/enken realized in four sentences representing the four

context plus boundary conditions.

### 4.3. Procedure

The procedure was the same as in Experiment 2. Again each stimulus was presented with four repetitions. This time, participants had to identify the target word as the existing word /z/enken, or as the non-word \*/s/enken, which was described to them as a ‘mispronunciation’. In line with the orthographic conventions of German, /z/enken was represented as *Senken* on the computer screen. The non-word /s/enken was represented as *Ssenken*, which is in line with the orthographic conventions for /s/ in word-medial position, where orthographic singleton ‘s’ typically corresponds to /z/.

### 4.4. Results

Figure 6 shows the average percentages of /z/-responses for the 20 steps in the four conditions. Again we analyzed the averaged percentages for each participant in a 3-way repeated measure ANOVA with Boundary, Context and Step as factors. All factors emerged as significant (Boundary:  $F(1, 19) = 138.690$ ,  $p < 0.001$ ; Context:  $F(1, 19) = 31.200$ ,  $p < 0.001$ , Step:  $F(19, 361) = 311.130$ ,  $p < 0.001$ ) as did their two-way interactions (Boundary by Context:  $F(1, 19) = 46.496$ ,  $p < 0.001$ ; Boundary by Step:  $F(19, 361) = 13.025$ ,  $p < 0.001$ , Context by Step:  $F(19, 361) = 6.20$ ,  $p < 0.001$ ) and the three-way interaction ( $F(19, 361) = 6.994$ ,  $p < 0.001$ ).

Figure 6 about here

In order to study the 3-way interaction in detail, we first analyzed the Boundary conditions separately. For the Phrase condition, we found main effects of Context ( $F(1, 19) = 38.717$ ,  $p < 0.001$ ) and Step ( $F(19, 361) = 193.004$ ,  $p < 0.001$ ) as

well as an interaction between these two predictors ( $F(19, 361) = 2.866, p < 0.001$ ). Separate analyses of all steps revealed that the Context effect, resulting in more /z/-responses in the assimilation context than in the non-assimilation context, was significant only for Steps 4, 5, 7 – 12, that is, for the ambiguous middle part of the continuum (pairwise comparisons:  $p < 0.05$ ). The Word condition also demonstrated main effects of Context ( $F(1,19) = 115.449, p < 0.001$ ) and Step ( $F(19, 361) = 207.430, p < 0.001$ ), and a significant interaction between these two ( $F(19, 361) = 15.146, p < 0.001$ ). The Context effect was present for Steps 4 – 16 (pairwise comparisons;  $p < 0.05$ ), and thus covered more of the voiced part of the continuum than in the Phrase condition.

In both prosodic conditions, more /z/-responses were obtained in assimilation context than in non-assimilation context, as illustrated in Figure 6. These results present evidence for compensation for assimilation. Apparently, compensation for assimilation also occurs for fricatives without lexically distinctive voiceless counterparts.

We also analyzed the assimilation and non-assimilation contexts separately. For non-assimilation context, we found main effects of Boundary ( $F(1, 19) = 92.175, p < 0.001$ ) and Step ( $F(19, 361) = 193.062, p < 0.001$ ) and a significant interaction ( $F(19, 361) = 11.290, p < 0.001$ ). The effect of Boundary was restricted to Steps 6, 8-16 ( $p < 0.05$ ). For these steps, there were more /z/-responses after a Phrase boundary than after a Word boundary. This pattern of results mirrors the production data, which showed that in /ə/-context, /z/ is more likely to be completely voiced after a Word boundary than after a Phrase boundary (Kuzla et al. in press). This suggests that listeners also compensate for aerodynamically induced devoicing, which may result from increase in duration due to prosodic strengthening.

For assimilation (/t/-) context, only Step was significant ( $F(19,361) = 176.679,$

$p < 0.001$ ). In other words, we did not observe an effect of prosodic structure for the assimilation context. Our perception data thus do not mirror the production data, which showed more glottal vibration after a Phrase boundary than after a Word boundary in /t/-context. We will return to this difference between /ə/-context and /t/-context in the General Discussion.

## **5. General Discussion**

This study investigated the role of segmental context and prosodic structure in the identification of Standard German word-initial fricatives. In Standard German, the lenis fricatives /v/ and /z/ can be devoiced after voiceless obstruents (e.g., /t/). Kuzla et al. (in press) showed that the amount of devoicing depends on the prosodic boundary separating the voiceless obstruent and the lenis fricative, as there is more devoicing across smaller prosodic boundaries. Since glottal vibration is a main cue to the fortis-lenis distinction in fricatives, devoicing may affect the identification of fricatives. In the present paper, we investigated whether listeners compensate for the devoicing. We examined whether their identification of fricatives as fortis or lenis is affected by segmental context and by the prosodic structure of the sentence. We conducted three identification experiments in which we varied the amount of glottal vibration of /v/ (Experiment 1 and 2) and /z/ (Experiment 3), preceded by a phrase boundary or a word boundary, and /t/ (assimilation context) or /ə/ (non-assimilation context).

All three experiments showed that listeners compensate for assimilatory devoicing. Participants identified ambiguous fricatives which were only partly produced with glottal vibration more often as lenis after /t/ than after /ə/. This finding extends the existing evidence for perceptual compensation for assimilation in two



important respects. So far, almost all research on compensation for assimilation has focused on regressive assimilation processes (except for Coenen et al. 2001), while our experiments show compensation for progressive assimilation. Furthermore, most previous studies examined the assimilation of place and manner of articulation (e.g., Gow 2001, 2002; Nolan 1992; Mitterer and Blomert 2003, Mitterer, Csépe, and Blomert 2006), whereas only few studies focused on voice assimilation (Coenen et al 2001; Darcy 2002, 2006; Gow and Im 2004; Snoeren et al 2006).

In Standard German, the /v-f/ contrast is lexically functional in word-initial position, given minimal word pairs such as /vɛldə/ - /fɛldə/ *Wälder - Felder* 'forests' – 'fields', /vazə/ - /fazə/ *Vase- Phase* 'vase' – 'phase', and /vɪʃən/ - /fɪʃən/ *wischen- fischen* 'to wipe' – 'to fish'. The /z-s/ contrast, on the contrary, is only functional in word-medial position, as there are no Standard German words starting with /s/. Notwithstanding this functional difference, listeners compensated for assimilatory devoicing of /z/, as they did for /v/. This finding suggests that compensation for voice assimilation is rooted in early auditory processes, as has been proposed for compensation for other assimilation processes (e.g., Mitterer, Csépe, and Blomert 2006). An alternative explanation, however, is also possible. Children might learn to compensate first for the devoicing of fricatives for which the fortis-lenis distinction is lexically relevant. This compensation may then generalize from functional to non-functional contrasts, for which compensation is unnecessary. Obviously, further research is required in this area.

This study is the first to investigate the role of prosodic structure in compensation for assimilation. Experiment 1 did not provide support for a role of prosody, probably because of ceiling effects. Experiment 2, in contrast, demonstrated that, as predicted, listeners interpret partially devoiced and therefore ambiguous fricatives in assimilation context more often as lenis after a word boundary than after

a phrase boundary.

The most important difference in experimental design between Experiment 1 and Experiment 2 is the duration of the fricatives. In Experiment 1, the duration was kept constant within a prosodic condition and corresponded to the relatively short duration of the lenis fricative. This short duration may account for the overall /v/-bias observed in this experiment: The percentages of /v/-responses in assimilation context ranged only from 77% to 91%. This small range may explain the absence of a prosodic structure effect. We chose to vary glottal vibration only such that our stimuli mirrored the production data (Kuzla et al. in press) most closely.

In Experiment 2, the two endpoints of the continuum were natural tokens of /v/ and /f/, and the durations of the intermediate steps covaried with the amount of vocal fold vibration. Hence the continuum reflected the natural fortis-lenis contrast in a better way. As a consequence, the percentages of /v/-responses in assimilation context ranged from 0% -100%. More importantly, we did observe an effect of prosodic structure.

In natural speech, fricatives tend to be longer after phrase boundaries than after word boundaries due to prosodic strengthening. In Experiment 2, the fricatives had the same duration in the two prosodic conditions, since we used only one continuum generated from phrase-initial fricatives. This implies that the fricatives following a word boundary were relatively long, which might have biased listeners' perception towards fortis. Nevertheless, we found that the partially devoiced fricatives were more often identified as lenis after a word boundary than after a phrase boundary, despite their relative longer duration. Hence, the attested effect of prosodic structure is an effect on compensation for assimilation.

Experiment 3 showed an effect of prosodic structure on the identification of /z/ in the non-assimilation context (i.e., after /ə/), which we did not observe in the /v/-

identification experiment: Listeners identified devoiced sounds from the middle of the continuum more often as lenis in the Phrase condition than in the Word condition. In other words, in the Word condition more glottal vibration was required to yield a /z/-percept. We propose the following explanation for this finding. Because of prosodic strengthening, phrase-initial fricatives are longer than word-initial fricatives. Vocal fold vibration is presumably more difficult to maintain throughout a longer fricative for aerodynamic reasons. This is assumed to be especially the case for alveolar fricatives, since they are produced with a smaller oral cavity than labiodental fricatives (Stevens et al. 1992). Hence, in the non-assimilation context partial devoicing is more common for phrase-initial /z/ than for word-initial /z/, or for both phrase- and word-initial /v/ (Kuzla et al. in press). Our results suggest that listeners compensate for this tendency found in speech production. If this explanation is correct, it suggests that listeners are surprisingly sensitive to the systematic fine phonetic variation in the speech signal (cf. Andruski, Blumstein and Burton 1994; Salverda, Dahan, and McQueen 2003; Kemps, Ernestus, Schreuder, and Baayen 2005).

In the assimilation context /t/, we did not observe listeners' sensitivity to aerodynamically-conditioned devoicing, which would have led to more /z/ - responses in the phrase condition. One possible explanation is that this type of compensation for devoicing is cancelled out by compensation for assimilatory devoicing, which results in more /z/-responses after a Word boundary.

An alternative explanation for the observed effect of prosodic structure on the identification of devoiced /z/ after /ə/, i.e., in our non-assimilation condition, is that German listeners are used to distinguishing between /z/ and /s/ in intervocalic word-medial position (e.g., rei/z/en 'travel' versus rei/s/en 'tear'). In our assimilation context, on the other hand, the fricative occurred after /t/ where the /z/-/s/-contrast is

never lexically functional. Accordingly, we did not observe an effect of prosodic structure on compensation for assimilatory devoicing of /z/.

Even though the offline nature of the categorization task does not allow us to make a strong claim for an online influence of the prosodic structure on phonemic decoding, the prosodic effects on compensation for assimilation observed in the present study suggest that listeners analyze the prosodic structure of spoken utterances as soon as the acoustic signal comes in. Similar conclusions have been drawn by Salverda, Dahan, and McQueen (2003). In an eye-tracking study, these authors demonstrated that a monosyllabic word embedded at the onset of a polysyllabic target word (e.g., *ham* in *hamster*) was activated more if it was longer and thus signaled an upcoming prosodic word boundary. Prosody also affects syntactic processing (e.g., Carlson, Clifton, and Frazier 2001, Kjelgaard and Speer 1999), which also implies that prosodic representations are computed while the incoming signal is being analyzed. The present study is the first to demonstrate that the computed prosodic representations may affect phonemic decoding: A partially devoiced fricative may be interpreted as either lenis or fortis depending on the type of preceding prosodic boundary.

How can current accounts of compensation of assimilation explain an influence of prosodic boundary strength on compensation? This is relatively unproblematic for Gaskell's (2003) probabilistic inference model, which could easily be extended to take prosodic information into account. The degree to which phonological inference is applied would depend on the prosodic boundary strength. Accounting for a prosodic effect is less straightforward in Gow's feature-parsing model (2002), which assumes that compensation occurs on the basis of a grouping of phonetic features: The observed devoicing in a fricative is grouped with the preceding voiceless stop in a [hatyɛldə] sequence. A prosodic influence would have to be

indirect, by influencing the physical "distance" of the features in the input, which in turn influences the grouping process. Finally, the auditory account proposed by Mitterer and colleagues (Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, and Blomert 2006) can easily explain the effect of prosody on compensation for assimilation. Mitterer, Csépe, and Blomert (2006) distinguish a general-auditory context effect, which would make the devoicing in [hatyeldə] difficult to perceive after /t/, and a language-specific bias towards perceiving phonemes which can undergo assimilation. A prosodic influence is easy to envision on the second, language-specific, process.

This paper combines the lines of research on compensation for assimilation and on the influence of prosodic structure on the fine phonetic detail of speech sounds. So far, many studies have been dedicated to compensation for assimilation, but none of them have taken into account that assimilation is constrained by prosodic structure. Our data show that prosody affects compensation for assimilatory devoicing, and that listeners adjust their phoneme categories to both segmental context and prosodic structure.

## **Acknowledgements**

We thank Jonathan Harrington and Ernst Dombrowski for providing testing facilities at Kiel University. This research was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

## References

- Andruski, Jean E., Blumstein, Sheila E., and Burton, Martha 1994 The effect of subphonetic differences on lexical access. *Cognition* 52, 163–187.
- Carlson, Katy, Clifton, Charles, Jr., and Frazier, Lyn 2001 Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45, 58--81.
- Cho, Taehong and Keating, Patricia A. 2001 Articulatory and acoustic studies on domain initial strengthening in Korean. *Journal of Phonetics* 29, 155--190.
- Cho, Taehong and McQueen, James M. 2005 Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33, 121--157.
- Cho, Taehong, McQueen, James M., and Cox, Ethan A. 2007 Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35, 210--243.
- Coenen, Else, Zwitserlood, Pienie and Bölte, Jens 2001 Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* 16, 535--564.
- Darcy, Isabelle 2002 Online processing of phonological variation in speech comprehension: The case of assimilation. In Hawkins, S. and Nguyen, N. (Eds.), *Temporal integration in the perception of speech*. Proceedings ISCA, Aix-en-Provence, April 8-10, 38.
- Darcy, Isabelle, Peperkamp, Sharon, and Dupoux, Emmanuel (to appear) Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners. In Cole, J. and Hualde, J. (Eds.), *Papers in Laboratory Phonology IX*.
- Ernestus, Mirjam, Lahey, Mybeth, Verhees, Femke and Baayen, R. Harald 2006 Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America* 120, 1040—1051.

Fougeron, Cécile 2001 Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29, 109--135.

Fougeron, Cécile and Keating, Patricia A. 1997 Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101, 3728--3740.

Gaskell, Gareth M. 2003 Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31, 447--463.

Gaskell, Gareth M. and Marslen-Wilson, William D. 1997 Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24, 380—396.

Gow, David W., Jr. 2001 Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45, 133--159.

Gow, David W., Jr. 2002 Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28, 163--179.

Gow, David W., Jr. and Im, Aaron M. 2004 A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* 51, 279—296.

Keating, Patricia A., Cho, Taehong, Fougeron, Cécile and Hsu, Chai-Shune 2003 Domain-initial articulatory strengthening in four languages. In Local, J., Ogden, R. and Temple, R. (Eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press. 145--163.

Kemps, Rachèl J., Ernestus, Mirjam, Schreuder, Robert and Baayen, R. Harald 2005 Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory and Cognition* 33, 430--446.

Kjelgaard, Margaret M. and Speer, Shari R. 1999 Prosodic facilitation and



interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40, 153--194.

Kohler, Klaus J. 1990 Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In Hardcastle, W. J. and Marchal, A. (Eds.) *Speech production and Speech Modelling*. Dordrecht: Kluwer. 69--92.

Kohler, Klaus J. 1995 *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

Kuzla, Claudia, Cho, Taehong and Ernestus, Mirjam in press Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics*.

Jessen, Michael 1998 *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.

Jun, Sun-Ah 1998 The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15, 189-226.

Mitterer, Holger 2003 Understanding “garden bench”. Studies on the perception of assimilated word forms. PhD Dissertation, Maastricht.

Mitterer, Holger and Blomert, Leo 2003 Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics* 65, 575--590.

Mitterer, Holger, Csépe, Valéria and Blomert, Leo 2006 The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology* 59, 1305--1334.

Mitterer, Holger, Csépe, Valéria., Honbolygo, Ferenc, and Blomert, Leo 2006 The recognition of assimilated word forms does not depend on specific language experience. *Cognitive Science* 30, 451--479.

Nolan, Francis 1992 The descriptive role of segments: Evidence from assimilation. In Docherty, G. J. and Ladd, D. R. (Eds.) *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press. 261-- 280.

Norris, Dennis 1994 Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52, 189—234.

Repp, Bruno. H., & Liberman, Alvin. M. 1987 Phonetic categories are flexible. In Harnad, S. (Ed.), *Categorical Perception: The groundwork of cognition*. Cambridge, Mass.: Cambridge University Press. 89—112.

Salverda, Anne Pier., Dahan, Delphine and McQueen, James M. 2003 The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90, 51--89.

Snoeren, Natalie D., Hall↓, Pierre A. and Segui, Juan 2006 A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* 34, 241—268.

Stevens, Kenneth N., Blumstein, Sheila E., Glicksman, Laura, Burton, Martha and Kurowski, Kathleen 1992 Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America* 91, 2979--3000.

Table 1: Speech materials for Experiment 1 and 2. The /v/-initial target word was “Wälder” (underscored). A contrastive accent on another word, indicated in bold, induced deaccentuation of the target word.

---

Phrase Boundary, Assimilation Context

---

Weil sie vorhat, Wälder und **Wiesen** zu malen, fährt sie aufs Land.

Because she plans, forests and meadows to paint, goes she to-the countryside.

*‘Since she plans to paint forests and meadows, she goes to the countryside.’*

---

Phrase Boundary, Non-assimilation Context

---

Weil sie vorhatte, Wälder und **Wiesen** zu malen, fuhr sie aufs Land.

Because she planned, forests and meadows to paint, went she to-the countryside.

*‘Since she planned to paint forests and meadows, she went to the countryside.’*

---

Word Boundary, Assimilation Context

---

**Anna** hat Wälder und Wiesen gemalt.

Anna has forests and meadows painted

*‘Anna has painted forests and meadows.’*

---

Word Boundary, Non-assimilation Context

---

**Anna** hatte Wälder und Wiesen gemalt.

Anna had forests and meadows painted

*‘Anna had painted forests and meadows.’*

---

Table 2: Additional speech materials for Experiment 2. The /f/-initial target word was “Felder”(underscored). A contrastive accent on another word, indicated in bold, induced deaccentuation of the target word.

---

Phrase Boundary, Assimilation Context

---

Weil sie vorhat, Felder und **Wiesen** zu malen, fährt sie aufs Land.

Because she plans, fields and meadows to paint, goes she to-the countryside.

*‘Since she plans to paint fields and meadows, she goes to the countryside.’*

---

Phrase Boundary, Non-assimilation Context

---

Weil sie vorhatte, Felder und **Wiesen** zu malen, fuhr sie aufs Land.

Because she planned, fields and meadows to paint, went she to-the countryside.

*‘Since she planned to paint fields and meadows, she went to the countryside.’*

---

Word Boundary, Assimilation Context

---

**Anna** hat Felder und Wiesen gemalt.

Anna has fields and meadows painted

*‘Anna has painted fields and meadows.’*

---

Word Boundary, Non-assimilation Context

---

**Anna** hatte Felder und Wiesen gemalt.

Anna had fields and meadows painted

*‘Anna had painted fields and meadows.’*

---

Table 3: Speech materials for Experiment 3. The /z/-initial target word was “Senken”(underscored). A contrastive accent on another word, as indicated in bold, induced deaccentuation of the target word. A second set of sentences was recorded where “Senken” was replaced by the /s/-initial non-word “Ssenken”.

---

Phrase Boundary, Assimilation Context

---

Weil sie vorhat, Senken und **Hügel** zu malen, fährt sie nach Holstein.

Because she plans, hollows and hills to paint, goes she to Holstein.

*‘Since she plans to paint hollows and hills, she goes to Holstein.’*

---

Phrase Boundary, Non-assimilation Context

---

Weil sie vorhatte, Senken und **Hügel** zu malen, fuhr sie nach Holstein.

Because she planned, hollows and hills to paint, went she to Holstein.

*‘Since she planned to paint hollows and hills, she went to Holstein.’*

---

Word Boundary, Assimilation Context

---

**Anna** hat Senken und Hügel gemalt.

Anna has hollows and hills painted

*‘Anna has painted hollows and hills.’*

---

Word Boundary, Non-assimilation Context

---

**Anna** hatte Senken und Hügel gemalt.

Anna had hollows and hills painted

*‘Anna had painted hollows and hills.’*

---

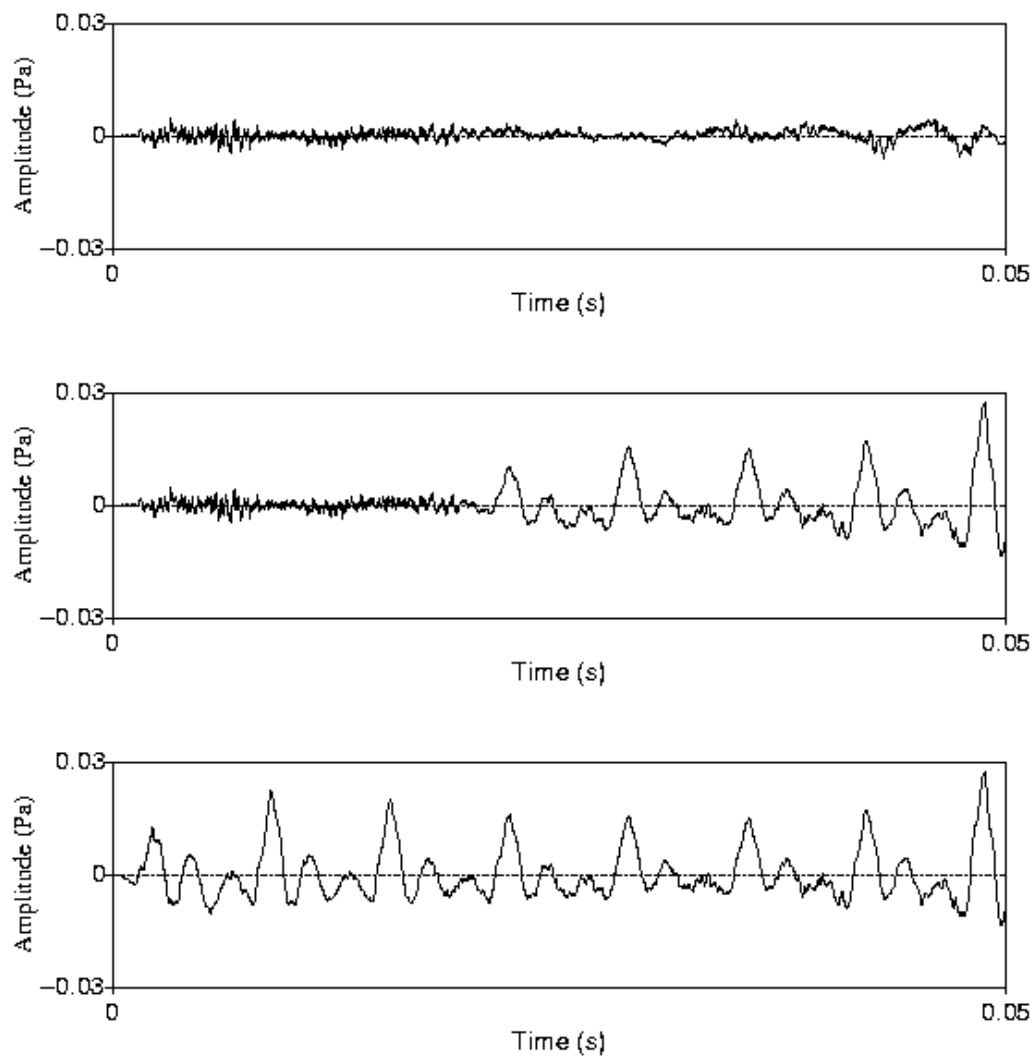


Figure 1: Acoustic waveform of the steps 1, 4 and 7 from the /y - v/ continuum for the prosodic word condition in Experiment 1

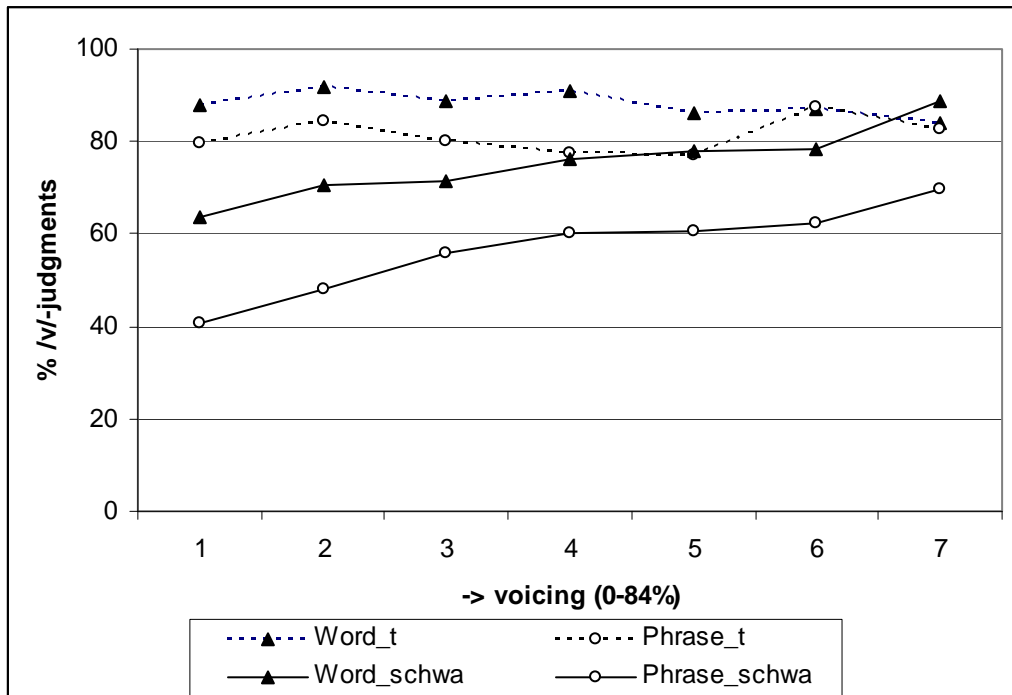


Figure 2: Percentages of /v/-judgments across the /y - v/ continua in the four prosodic+segmental context conditions in Experiment 1

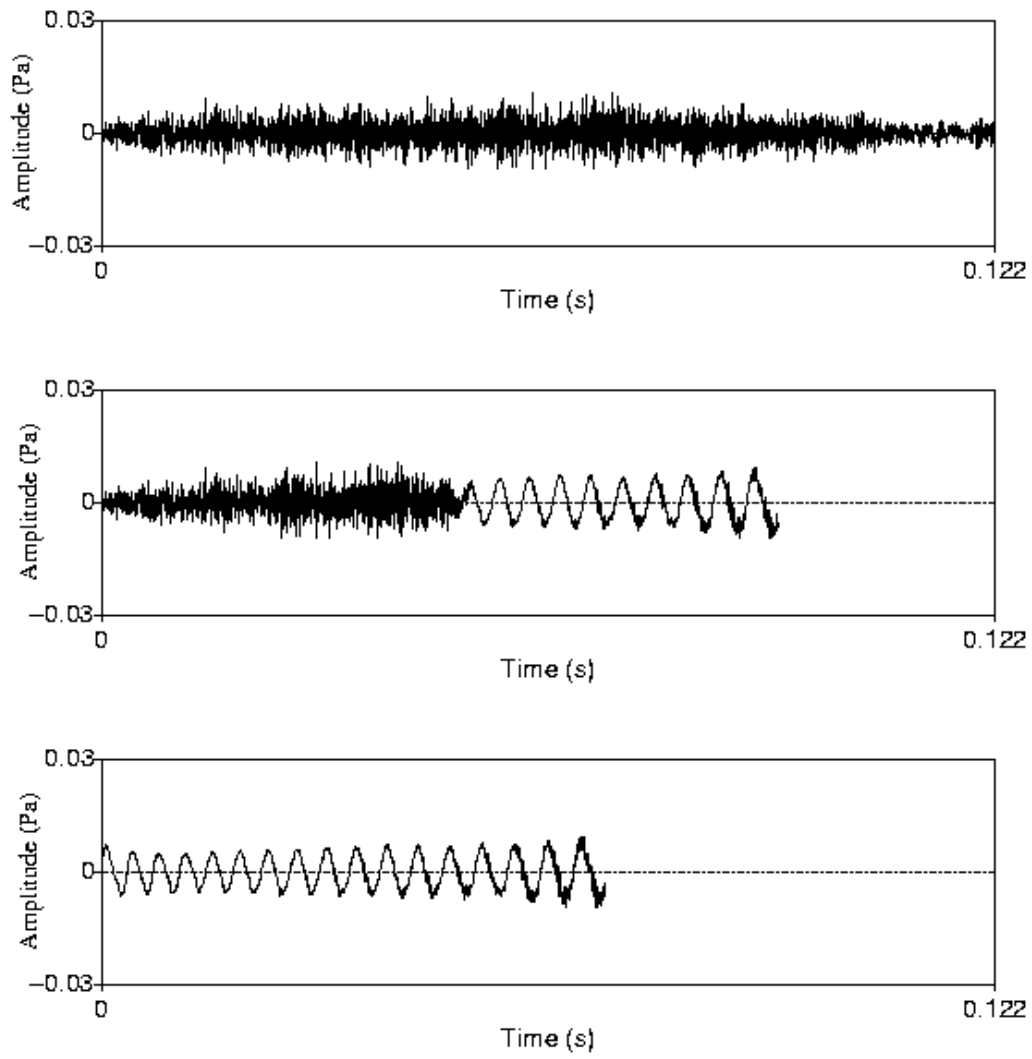


Figure 3: Acoustic waveforms of the steps 1, 10 and 20 from the /f - v/ continuum in Experiment 2



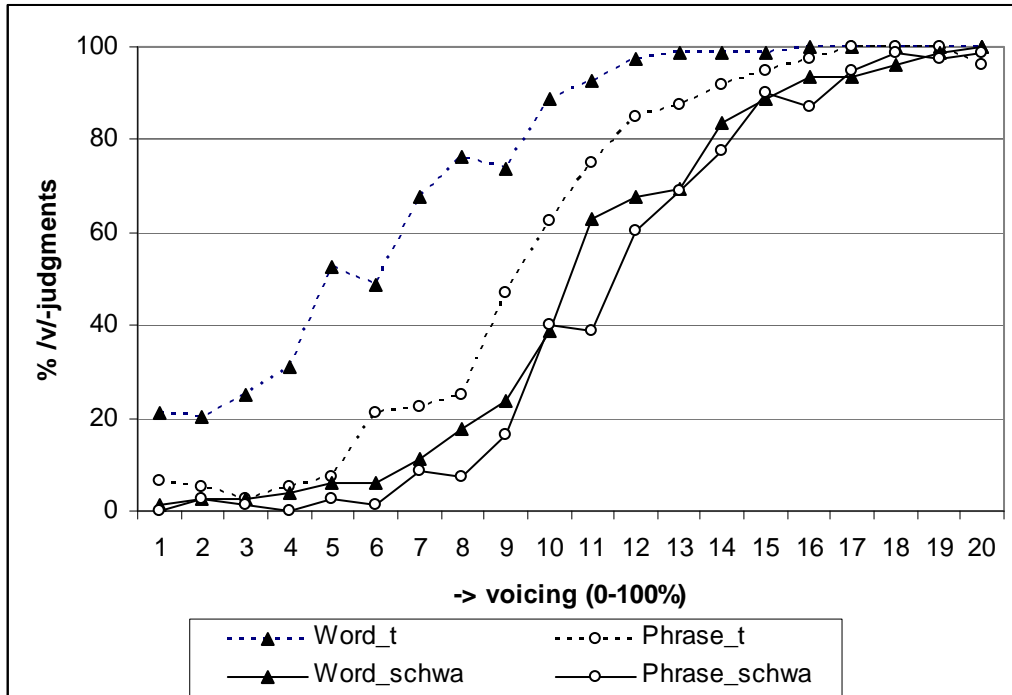


Figure 4: Percentages of /v/-judgments across the /f - v/-continuum in the four prosodic+segmental context conditions in Experiment 2

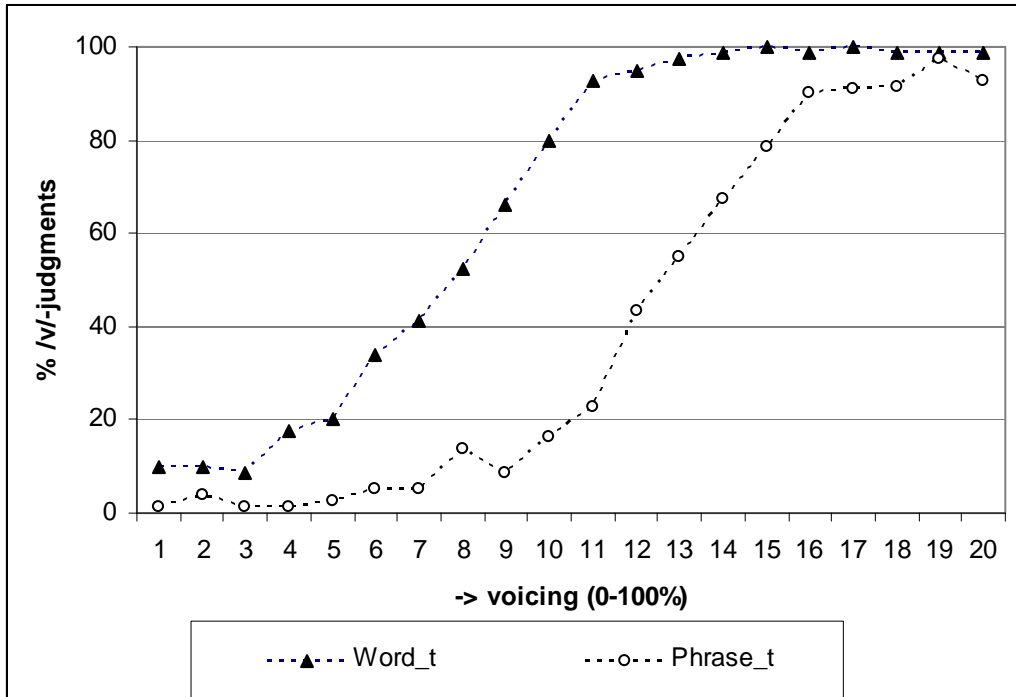


Figure 5: Percentages of /v/-judgments across the /f - v/-continuum in two prosodic conditions in the assimilation context, carrier sentences containing “Felder” in Experiment 2

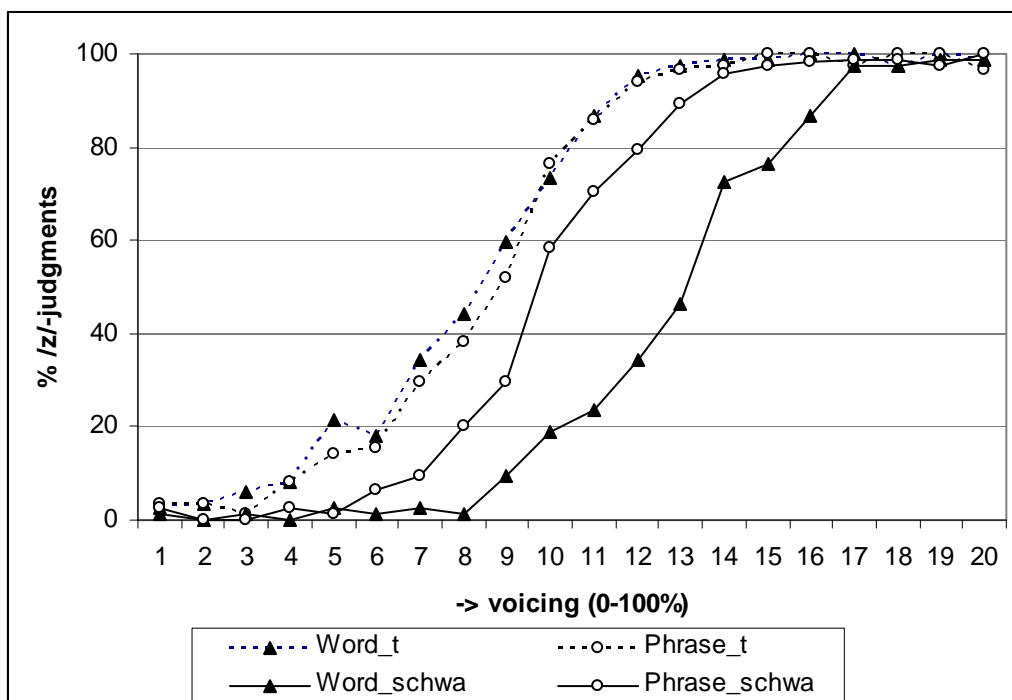


Figure 6: Percentages of /z/-judgments across the /s - z/-continuum in the four prosodic+segmental context conditions in Experiment 3