Prosodic structure

in speech production and perception

ISBN: 978-90-76203-33-1

COVER DESIGN: Linda van den Akker, Inge Doehring

COVER ILLUSTRATION: Klaus Langenfeld

Printed and bound by Ponsen & Looijen bv, Wageningen

Prosodic Structure

in Speech Production and Perception

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann
volgens besluit van het college van decanen
in het openbaar te verdedigen
op dinsdag 14 april 2009
om 13:30 uur precies

door

Claudia Kuzla

geboren op 28 maart 1973
te Neustadt/Holstein, Duitsland

PROMOTOR: Mevrouw prof. dr. A. Cutler

COPROMOTOR: Mevrouw dr. M. Ernestus (MPI, Nijmegen)

MANUSCRIPTCOMMISSIE:

Prof. dr. C. Gussenhoven

Mevrouw prof. dr. P. Zwitserlood (Universität Münster, Germany)

Mevrouw dr. C. Fougeron (CNRS - Université Paris 3, France)

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

In memory of my mother

Obstacles are those frightful things you see when you take your eyes off your goal.

- *Henry Ford*

Acknowledgements

When starting a PhD project at the MPI, I anticipated that writing a thesis would require a fair amount of personal commitment. I had no idea, however, that it would also need an entire social network. Finally, I have the opportunity to thank all those who made it possible.

First, I would like to express my gratitude to my promotor Anne Cutler. Thanking Anne first and foremost is not just good manners: Without her readiness to accept an ‘odd’ PhD student like me, my academic career might have ended before even begun. Thank you, Anne, for your confidence and encouragement in any circumstances, for the opportunity to participate in so many summer schools, courses, workshops and international conferences, for your quick and clear comments on my writing, and for the long-term guidance of my research. I greatly benefitted from the unique atmosphere in the Comprehension Group, and would like to thank everyone who has been part of it over the years.

Mirjam Ernestus has been involved in the planning and realization of my PhD research from the beginning to the end, and I always appreciated her thoughtful advice. Step by step she walked me through the data jungle until I could stand on my own feet, and she always managed to make me see the forest and the field again when it seemed all trees and mud to me. Mirjam, how can I thank you?

The cooperation with Taehong Cho taught me many a valuable lesson. I would like to thank you, Taehong, for putting me on the prosodic structure track and for an exciting and high-speeded first phase of this PhD project - just remember the unforgotten moments of harmony after my first presentations in Barcelona and Urbana-Champaign!

After Taehong had left for Korea, Holger Mitterer became my co-advisor in the second half of the project. Holger, your contribution can hardly be overestimated. Thanks for being an entertaining lunch table lecturer (e.g., on SciFi literature), the local PRAAT expert, a sophisticated Perl programmer, a critical reader, and a skilled experimental psychologist - I benefitted greatly from all of that.

Many thanks are due to Jonathan Harrington and Ernst Dombrowski, who hosted me three times at Kiel University and provided facilities for the data collection. Thanks to the lively research group at the IPDS Kiel, including in those days Tine, Michel, Benno, Thomas, Tamara, Oliver, Felicitas, Lasse, and Gilbert, it was always a pleasure to be there. And thanks to all my participants!

Rebecca Özdemir and Heidrun Bien were the voices in my perception experiments and patiently recorded all those funny sentences. Marianne Radel has labeled several thousands of boundary tones, and kindly hosted me in Hamburg.

At the MPI, special thanks are due to the student deans, the PhD representatives, the secretaries - in particular Rian Zondervan-, and to everybody from the Technical Group, the library, the administration, the reception and the canteen for doing their jobs as they did.

The PhD group as a whole was a constant source of inspiration and support. Frank and Ethan, my first office mates, showed me all the little secrets to get started. Anne Pier, Christiane and Martijn followed, and with Annelie 'Tuinvrouw' came plants and more feminine wall decorations to the office. Matthias took care for my stuff even after I had left. Petra and Kerstin stood by my side in the darkest hour. Mirjam was a caretaking PhDMum and a generous local host after several late night events in Nijmegen. Anita was always great company, especially in Düsseldorf and Barcelona, and Keren generously shared her experience with NESU, SPSS and all the other technical thingies I just couldn't grasp from reading the lovely manuals.

Anne Pier and Mirjam, thank you for the Dutch samenvatting.
Gertie und Adriana, vielen herzlichen Dank Euch, dass Ihr Eurer freundschaftlichen Unterstützung während der ganzen Zeit nun die Krone aufsetzt, indem Ihr meine Paranimpfen seid.

Klaus Langenfeld hat viel mehr beigetragen als die *Felder-Wälder*-Illustration auf dem Cover, denn ohne ihn wäre aus mir keine Linguistin geworden. Für die konstante Unterstützung in mancherlei Form in all den Jahren *gratias tibi ago, magister*.

Herzlichen Dank an Berna, aber auch an Anke und Martin sowie Tine und Anne, dass Ihr in Kleve und Kiel Euer Zuhause sowie manche Freuden und Leiden der akademischen Lebensform mit uns geteilt habt.

Was Gitte Büttner und ihre Familie getan haben, die Judith während der Jahre am Niederrhein ein zweites Zuhause boten, so dass ich ohne das spezifisch deutsche „Rabenmutterschuldgefühl“ meine Arbeit tun konnte, das weiß der Himmel. An dieser Stelle möchte ich auch allen anderen Klever Freunden danken, die mit Kinderbetreuung rund um die Uhr eingesprungen sind, wenn einmal wieder drei Tage vorm Kongress die Windpocken in der Kita ausgebrochen waren.

Zum Schluss möchte ich mich bei meiner Familie für die ungebrochene Zuversicht und Unterstützung bedanken. Bei wem ich mich bedanken soll, dass Judith im Windschatten dieser Arbeit so ein fröhliches, kooperatives Kind geworden ist, weiß ich auch nicht. Am besten bei ihr selber.

Contents

Introduction

Chapter 1

Encoding and decoding connected speech

People are usually able to understand each other without any noticeable difficulty, assuming that they speak the same language. However, the processes underlying the seemingly effortless verbal communications occupying a large part of our everyday lives are tremendously complex, and not yet completely understood.

Spoken language is an acoustic signal, a sequence of changes in air pressure. Speakers produce this signal by moving various parts of their vocal tracts, for instance, their vocal folds, tongues, jaws, and lips. These articulatory movements shape the air stream according to the meaning the speaker wants to convey. The listener recognizes portions of the signal which carry the meaning intended by the speaker, commonly called words. Words are the finite number of meaningful building blocks of language that combine into an infinite number of possible messages. Hence, words must be stored in memory, and an inevitable part of understanding spoken language is the identification of word forms in the acoustic signal, in order to access the meaning associated with them.

Whereas in written language words are separated by spaces, the acoustic speech signal is continuous. We typically do not insert pauses before and after each word in an utterance, which would sound very unnatural to most of us. On the other hand, interruptions and hesitations can occur in the middle of words, and this does not hinder us from putting the acoustic pieces together and still recognizing the word as a whole. If we know a language spoken to us, we cannot help perceiving the speech stream as a sequence of words. However, if we listen to a totally unfamiliar language, it becomes evident that this segmentation of the speech signal is the result of a cognitive process rather than a fact given in the acoustic substrate of language, because it turns out impossible to tell where words begin and end. Obviously, the continuous acoustic signal is mapped onto the lexical categories represented in our memory. This process requires knowledge about the sound form of the words in our language.

Crucially, the detailed acoustic form of one and the same word can be very different, depending for instance on the age, gender, health, and emotional state of the speaker, not to mention the particular regional or foreign accent, on the speech rate and speaking style, and on the acoustic characteristics of the environment. However, even the same speaker within the same situation will never produce two acoustically identical versions of the same word. Nevertheless, listeners are able to recognize words correctly.

Some sources of variation are located in the linguistic system, in particular in the way words are combined in connected speech. First, words typically do not occur in isolation. The Sanskrit term sandhi describes phonological changes in word forms which occur in word sequences. For instance, the English indefinite article has the form *a* if the following word starts with a consonant (*a cat*), but the form *an* if the following word starts with a vowel (*an animal).* The perhaps best-known and most common sandhi process is assimilation (see below). Second, there are different degrees of cohesion between words. Spoken language has a metrical and rhythmical structure, the so called prosodic structure, by which elements of speech are grouped into different-sized units, e.g., syllables into words, words into phrases, and eventually, smaller phrases into larger phrases and whole utterances (Figure1).

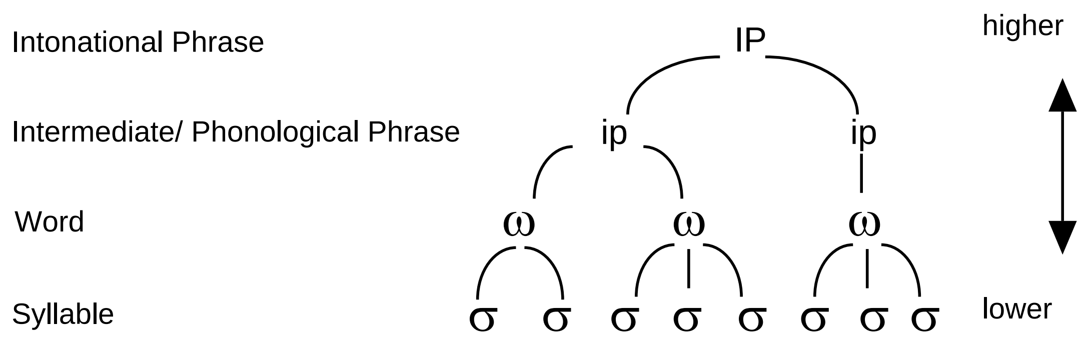


Figure 1: Prosodic Structure (adapted from Beckman and Pierrehumbert 1986)

While prosodic structure itself is manifested in the phonetic details of speech sounds, it also constrains sandhi processes. This thesis is concerned with systematic phonetic variation induced by the prosodic structure and how this prosodically-conditioned variation is relevant for spoken word recognition.

Minimal pairs

Words can differ from other words to various degrees. The word *back* is very different from the word *sunflower*, but quite similar to *pack*, *track*, *bat*, and *bag*. The difference between *back* and *pack* is indeed only in the initial, that between *back* and *bat* only in the final sound. If two sounds differentiate between words in a given language, they convey different phonemes. Two words which only differ in a single phoneme form minimal pairs. It is particularly interesting to study phonetic variation in minimal pairs, because another word could be recognized if the phonemes were not identified correctly. In cases where no other minimally different words exist, phonetic variation does not create lexical ambiguities; hence there can be more variation.

Coarticulation and assimilation

The realization of speech sounds is influenced by neighboring sounds. For instance, the /k/ in the German word *Kino* ‘cinema’ is produced at a more fronted position (towards the hard palate) than in *Kanu* ‘canoe’ (at the velum), because of the difference in the following vowel; accordingly, its acoustic release noise has different spectral characteristics. This type of variation, resulting in different variants of /k/, is known as coarticulation. Similarly, the place of articulation of the /n/ in *in* will be different in the phrases *in Paris* [impaˈriːs] ‘in Paris’, *in Nijmegen* [inˈnaemexə] ‘in Nijmegen’ and *in Genf* [iŋˈgɛɱf] ‘in Geneva’, potentially shifting from [m] to [n], to [ŋ] due to the following consonant. As /m/, /n/ and /ŋ/ are considered different phonemes in German, this process is called assimilation. In phonological terms, coarticulation produces variants (allophones) of the same phoneme whereas assimilation results in a different phoneme. Phonetic research, however, has shown that assimilation is often a graded process (e.g., Nolan 1992; Nolan, Holst and Kühnert 1996), such that articulatory and acoustic traces of the original segment are still present in the signal. Such fine-grained phonetic detail can be enough for listeners to resolve lexical ambiguities (e.g., Coenen, Zwitserlood and Bölte 2001; Gow 2002; Snoeren, Hallé and Segui 2006).

Prosodic structure and phonetic detail

Prosodic structure, as described above, is “a complex grammatical structure in its own right” (Beckman 1996; cf. Shattuck-Hufnagel and Turk 1996). What the prosodic structure of an utterance will be is determined by various grammatical and non-grammatical factors such as syntax, semantics, information structure, speech rate, and length of the utterance (Nespor and Vogel 1986, Jun 1993). In other words, depending on the speaker’s communicative intention, the same string of words can be uttered quite differently in terms of prosody, which can imply remarkable differences in meaning. Moreover, the forms of many words will change, because prosody governs various phonological and phonetic processes.

Within the framework of Prosodic Phonology (Nespor and Vogel 1986), prosodic constituents are, by definition, the domains of phonological processes. For example, consonat gemination (‘Raddoppiamento Sintattico’) in Italian, linking *r* in RP English, and liaison in French are sandhi phenomena which only occur within phonological phrases. Higher prosodic boundaries will block these processes.

In phonologically unchanged speech sounds too, phonetic details are conditioned by prosodic structure. In particular, phonetic patterns correlated with the edges of prosodic constituents, that is, boundary phenomena, have been repeatedly attested. For instance, in Korean, a voiceless lax stop consonant becomes voiced in intervocalic position, but the degree of voicing becomes gradiently larger as the prosodic boundary between the lax stop and the following vowel becomes progressively higher from a syllable to a phrase boundary (Jun 1995). Other studies (e.g., Cho and Keating 2001; Fougeron 2001; Keating, Cho, Fougeron and Hsu 2003; Jun 1998; Dilley, Shattuck-Hufnagel and Ostendorf 1996; Wightman, Shattuck-Hufnagel, Ostendorf and Price 1992) have also demonstrated the influence of prosodic boundary size on the adjacent speech sounds in other languages, for example in English, French, and Taiwanese. The most important effects of prosodic structure are domain-final lengthening, that is, a local slowing down of the speech rate before prosodic boundaries, and the so called domain-initial strengthening. Domain-initial strengthening implies the temporal and spatial expansion of articulatory gestures. One important consequence of prosodic strengthening is hence the reduction of coarticulation, which in turn suggests that word-initial phonemes should be recognized easier. On the other hand, gestural timing and strength of articulation play an important role in phonemic contrasts. The consequences of prosodic strengthening on phoneme identity are thus hard to predict, due to the complex mapping between articulation and acoustics. In addition, which acoustic characteristics serve as perceptual cues to phoneme contrasts differs from language to language.

Listeners' sensitivity to prosodically conditioned phonetic detail

Recent studies in speech comprehension have shown that listeners are sensitive to prosodically conditioned subphonemic variation in the signal. For example, subtle durational differences in phonemically identical sequences, which are supposed to reflect prosodic structure, affect word recognition. Salverda, Dahan and McQueen (2003) found that listeners make use of syllable duration to predict whether more syllables will follow within the same prosodic word or not. These durational cues distinguish, for example, between the longer monosyllabic word *ham* and the shorter syllable *ham-* embedded in *hamster*. Spinelli, McQueen and Cutler (2003) showed that subtle durational differences also resolve lexical ambiguities created by French liaison, a sandhi process that in some cases generates homophonous sequences at the phonemic level (e.g. *dernier rognon* 'last kidney' vs. *dernier oignon* 'last onion'). In exploring the effect of higher-level prosodic structure on speech comprehension, Cho, McQueen and Cox (2007) found that the first word in two-word sequences was recognized faster when the initial portion of the second word came from a phrase boundary context then when it was from a word boundary context (*…bus. Tickets…* vs. *…bus tickets…*), that is, when the initial segments of the second word had undergone domain-initial strengthening. Similar results were obtained by Christophe, Peperkamp, Pallier, Block and Mehler (2004), who attested that lexical competitor words (e.g., French *chagrin* 'pilgrim' in the sequence *...chat grimpait...* '[the] cat climbed') only get activated across word boundaries, but not across phrase boundaries. These findings demonstrate that the effects of prosodic structure on phonetic detail assist speech segmentation, and consequently word recognition.

This thesis

While the studies discussed in the previous section were concerned with the implications of prosodic structure for word segmentation, the major objective of the work presented here is to explore the interaction of prosodic strengthening and phonemic contrasts. As noted above, domain-initial articulatory strengthening can be manifested in various acoustic measures. Importantly, some of these acoustic characteristics could also be cues to phoneme distinctions. To investigate how prosody influences phoneme identity, I took into account both speech production and perception. Production studies are necessary to establish the effects of prosodic structure on phonetic detail, in particular because I investigate German, a language for which prosodic strengthening has not been examined so far. Perception experiments were carried out because production studies by themselves do not allow us to determine whether prosodically conditioned variation found in production is relevant for listeners in terms of phonemic identity, or merely irrelevant variation within categories.

In Chapter 2, I investigated the production of the plosive consonants /p, t, k/ versus /b, d, g/ at word onsets in different prosodic domains. The contrast between these plosives is commonly referred to as a Fortis-Lenis contrast (e.g., Kohler 1995), and it occurs in numerous minimal pairs such as *packen* ‘pack’ versus *backen* ‘bake’, *tanken* ‘load fuel’ versus *danken* ‘thank’, and *Karten* ‘cards’ versus *Garten* ‘garden’. Native speakers of German were asked to read sentences where such words occurred in different prosodic environments. In an acoustic analysis, I examined whether acoustic cues involved in the German fortis-lenis contrast are also subjected to prosodically-conditioned variation, and whether the process of prosodic strengthening can be interpreted as a general ‘fortition’ by gestural expansion, or as enhancement of distinctive features which set the two consonant classes apart.

Chapter 3 reports on prosodic strengthening as resistance to sandhi in fricatives. This class of obstruents also displays a phonological fortis-lenis contrast in German, which shows an interesting distributional asymmetry. While the labiodental fricatives /f/ and /v/ both occur in word initial position and differentiate between members of minimal pairs such as *Felder* ‘fields’ and ‘forests’, the alveolar fricatives /s/ and /z/ are only contrastive in word-medial position. Word-initially, only /z/ occurs in Standard German. Importantly, word-initial /v/ and

/z/ often undergo the sandhi process of progressive assimilatory devoicing, that is, they are produced without vocal fold vibration, if preceded by voiceless consonants. This sandhi process could make it harder to differentiate between *hat Felder* ‘has fields’ and *hat Wälder* ‘has forests’, since vocal fold vibration is assumed to be an important cue to the fortis-lenis contrast in fricatives. That no similar lexical ambiguity arises if /z/ is devoiced in *hat Sand* ‘has sand’, allows us to study the influence of lexical contrasts in a direct comparison with the *hat Wälder* case. Again, native speakers produced fricative-initial words in various prosodic conditions, and in contexts which licensed sandhi or did not do so. Acoustic analyses were carried out to examine the influence of prosody, licensing sandhi context, and the existence of a lexical contrast on the fortis-lenis cues in the fricatives.

The experiments in Chapter 4 and 5 further investigate the role of prosodic structure in the assimilatory devoicing of fricatives, now turning to perception. Chapter 4 reports a series of Phoneme Categorisation experiments (see, e.g., McQueen 1996 for a general description of this method). Native listeners heard sounds from acoustic continua between fortis and lenis fricative endpoints, or between voiced and devoiced lenis fricative endpoints in sentences which constituted various segmental and prosodic contexts. The task was to decide which word they had heard, for instance, *Felder* or *Wälder*. This experimental method gives insights into how listeners adapt their phoneme categories according to the segmental and prosodic contextual factors which influence the phonetic detail of fricatives in speech production.

There are, however, two objections which can be raised against phoneme categorisation as a method here. First, the large number of continuum steps and repetitions only allowed the testing of one minimal word pair per continuum. Second, phoneme categorisation is arguably a decision-based task which taps into late stages of speech processing. In order to meet these objectives, I also employed cross-modal priming to examine how listeners deal with prosodic structure and assimilatory devoicing of fricatives. The idea of this experimental paradigm, which is used in many variants (see, e.g., Tabossi 1996), is that listeners are faster in recognizing a word presented to them in the visual mode if they have just heard the same word (the prime) than if they have heard an unrelated prime. In other words, if in a certain context, an acoustic variant has been heard as *Wälder*, reactions to *Felder* would be slower than if the acoustic variant had been perceived as *Felder*. A great advantage of this method is that it not only permits, but indeed requires a large number of minimal word pairs, which gives insights in whether the effects of prosodic structure on compensation for assimilation can be generalized over the German lexicon. The cross-modal priming experiments are reported in Chapter 5. Finally, Chapter 6 provides a summary and a general discussion of the findings of this thesis.

References

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes* **11**, 17-67.

Beckman, M. E. and Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* **3**, 255-309.

Cho, T., McQueen, J.M. and Cox, E.A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* **35**, 210-243.

Christophe, A., Peperkamp, S., Pallier, C., Block, E. and Mehler, J. (2004). Phonological phrase boundaries constrain lexical access. I. Adult data. *Journal of Memory and Language* **51**, 523-547.

Coenen, E., Zwitserlood, P., and Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* **16**, 535-564.

Cole, J., Kim, H., Choi, H. and Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* **35**, 180-209.

Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* **24**, 423-444.

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

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

Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody*. PhD Dissertation. Ohio State University.

Jun, S.-A. (1998). The Accentual Phrase in the Korean Prosodic Hierarchy. *Phonology* **15**, 189-226.

Keating, P. A., Cho, T., Fougeron, C. and Hsu, C.-S. (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.

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

McQueen, J.M. (1996). Phonetic categorisation. *Language and Cognitive Processes* **11**, 655-664.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

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

Nolan, F., Holst, T., and Kühnert, B. (1996). Modelling [s] to [ʃ] accomodation in English. *Journal of Phonetics* **24**, 113-137.

Salverda, A., Dahan, D. and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* **90**, 51-89.

Shattuck-Hufnagel, S. and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* **25**, 193-247.

Snoeren, N.D., Hallé, P. A. and Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* **34**, 241-268.

Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes* **11**, 569-576.

Prosodic conditioning of phonetic detail in German plosives

Chapter 2

Kuzla, C. and Ernestus, M. (accepted). *Journal of Phonetics.*

Abstract

This study investigates the prosodic conditioning of phonetic details which are candidate cues to phonological contrasts. German /b, d, g, p, t, k/ were examined in three prosodic positions. Lenis plosives were produced with less glottal vibration at larger prosodic boundaries, whereas their VOT showed no effect of prosody. VOT of fortis plosives decreased at larger boundaries, as did their burst intensity maximum. Vowels following both fortis and lenis velars were shorter after larger boundaries. Closure duration, which did not contribute to the fortis/lenis contrast, was heavily affected by prosody. These results support neither of the hitherto proposed accounts of prosodic strengthening (Uniform Strengthening and Feature Enhancement). We propose a different account, stating that the phonological identity of speech sounds remains stable not only within, but also across prosodic positions (contrast-over-prosody hypothesis). Domain-initial strengthening does not neutralize the contrast between prosodically weak fortis and strong lenis plosives.

Introduction

Part of the variation in the speech signal is induced by the prosodic structure of the utterance. Prosodic structure is the hierarchical organization of spoken language into constituents such as phrases, prosodic words, syllables etc (e.g., Selkirk 1984, 1986; Nespor and Vogel 1986; see Shattuck-Hufnagel and Turk 1996 for an overview). These constituents, or prosodic domains, are embedded such that a larger prosodic domain comprises one or more next-lower constituents (i.e., a prosodic word consists of syllables, a phrase of prosodic words, and a major phrase or utterance of smaller phrases). Phonetic research in various languages has shown that speech sounds are articulated more strongly at the beginning of prosodic domains than in domain-medial position (e.g., Cho and Keating 2001, 2007; Cho 2005; Fougeron 2001; Fougeron and Keating 1997; Jun 1998; Keating, Cho, Fougeron and Hsu 2003; Tabain 2003a, b). This domain-initial prosodic strengthening is cumulative within the prosodic hierarchy, such that word-initial segments which are also in initial position of a phrase are stronger than word-initial, but phrase-medial segments. In English, for instance, plosives are mostly produced with longer closures, and often also with longer VOT (Voice Onset Time), after a phrase boundary than after just a word boundary (Cho and Keating 2007; Choi 2003; Fougeron and Keating 1997; Keating 1984; Pierrehumbert and Talkin 1992).

Importantly, some of the acoustic characteristics affected by prosodic strengthening are also cues to phonological contrasts. For instance, in both English and German, fortis[1] plosives are known to be produced with longer closures, longer VOT, stronger release noises and shorter preceding and following vowels than lenis plosives, while the presence and importance of these cues varies with segmental environment and position in the word (e.g., Allen and Miller 1999; Crystal and House 1988; Fischer-Jørgensen 1976; Jessen 1998; Lisker 1986; Kohler 1984; Peterson and Lehiste 1960). This raises the question of how prosodic structure and phonological contrast interact in determining the fine phonetic details of fortis

[1] Since [voice] is not a distinctive feature in German plosives (Jessen 2001), we refer to the phonological opposition between /p, t, k/ and /b, d, g/ as the fortis-lenis contrast, and reserve the terms “voiced”and “voicing” for *phonetically* voiced speech sounds, i.e., those produced with glottal vibration.

and lenis plosives. How do fortis and lenis plosives differ from each other in different prosodic positions? What happens, for instance, to the different cues to a /b/ if it undergoes prosodic strengthening? Is a prosodically strong /b/ distinct from a prosodically weak /p/?

The so called Uniform Strengthening Account (Cole, Kim, Choi and Hasegawa-Johnson 2007) predicts that all plosives are more fortis-like in stronger prosodic positions. Due to the expansion of articulatory gestures, both /b/ and /p/ are produced with a longer closure, with a longer VOT, and a greater intensity of the release noise at higher prosodic boundaries. As a consequence, lenis plosives at low prosodic boundaries differ maximally from fortis plosives at high prosodic boundaries. This hypothesis is supported by the overall strengthening patterns of English plosives. For instance, closure duration is longer for fortis than for lenis plosives, and all plosives are longer in prosodically stronger positions. Voice Onset Time, which is longer for fortis than for lenis plosives, is also longer in prosodically stronger plosives (Cho and Keating 2007; Pierrehumbert and Talkin 1992; but see Choi 2003; Cole et al. 2007, who did not find a prosodic position effect). If this uniform strengthening effect were universal, it would also explain why sound changes such as consonant lenition and deletion occur first in prosodically weak positions, as noted by Fougeron (2001).

The Feature Enhancement Account (e.g., Cho 2005; Cho and Jun 2000; de Jong 1995), predicts that the contrast between fortis and lenis plosives is more pronounced in prosodically stronger positions. To our knowledge, there is no evidence for paradigmatic contrast enhancement in the domain-initial position. However, an interesting observation by Cho and McQueen (2005), who studied prosodic effects on Dutch obstruents, has been interpreted by these authors as domain-initial enhancement of phonetic features. Cho and McQueen report shorter VOT for Dutch fortis plosives at larger prosodic boundaries, in contrast to what has been found for English. They relate their findings to an important difference between English and Dutch in the partitioning of the voice onset time dimension. In English, both fortis and lenis plosives typically display positive VOT, which is substantially longer for the voiceless plosives. In Dutch, lenis plosives are produced with prevoicing (i.e., negative VOT), while fortis plosives show positive VOT, although shorter than in English (Slis and Cohen 1969; van Alphen and Smits 2004). Cho and McQueen argue that English fortis plosives are characterized by the phonetic feature {+spread glottis}, whereas Dutch fortis plosives are {-spread glottis}. The finding that fortis plosives in stronger prosodic positions are produced with longer VOT in English, but with shorter VOT in Dutch can thus be explained as enhancement of language-specific phonetic features.

Recently, it has been recognized (Cho and Keating 2007) that there may exist important differences between the accented position and the domain-initial position. Comparison of accented and unaccented words suggests that cues to paradigmatic contrasts, such as in *bear – pear,* are enhanced under phrasal accentuation. Domain-initial strengthening, on the other hand, would mainly enhance syntagmatic contrasts, that is, the contrasts between neighboring segments, in particular those between consonants (C) and vowels (V) in CV onsets (Beckman, Edwards and Fletcher 1992; Cho 2004, 2005; Cho and Keating 2007; Cole et al. 2007; Farnetani and Vayra 1996). Obviously, further research is necessary to shed more light on the question of how domain-initial strengthening affects paradigmatic contrasts.

The current study investigates the interplay of prosody and the phonological fortis-lenis contrast in German, another West Germanic language, closely related with both English and Dutch. The cues to the contrast in German are different from those in both English and Dutch, and therefore allow us to evaluate the Uniform Strengthening and the Feature Enhancement accounts of prosodic strengthening.

Depending on the segmental environment and position in the word, there are numerous cues to the fortis-lenis contrast in German (Jessen 1998, 2001; Kohler 1984, 1995). The most important cue or ‘basic correlate’ according to Jessen (1998, 2001), is VOT. As in English, German fortis plosives are produced with considerable aspiration, while the lenis plosives also show positive, but shorter VOT (Fischer-Jørgensen 1976; Jessen 1998). In terms of Cho and McQueen (2005), German fortis plosives are therefore characterized by {+spread glottis}, which is also supported by glottographic data (Butcher 1977; Fuchs 2005; see also Jessen and Ringen 2002).

One of the non-basic correlates of the fortis-lenis contrast is the duration of the preceding vowel. Even though this cue is not as important as in English, in German vowels tend to be longer before lenis plosives as well (Kohler 1979). An additional cue is glottal vibration during the closure. Although glottal vibration (‘voicing’) is not a necessary correlate of lenis plosives in general, it often occurs in intervocalic position (Fischer-Jørgensen 1976; Jessen 1998). Another candidate cue is the intensity of the burst, which has been claimed to be greater for fortis than for lenis plosives (Kohler 1977, 1995).

Other reported cues include the duration of the closure, which is longer for fortis plosives at least in word-medial position, and the fundamental frequency onset as well as the first formant onset in the following vowel, which are both higher after fortes (Jessen 1998; Kohler 1982).

The present study investigates word-initial plosives preceded and followed by a vowel. We opted for plosives in CV onsets, since most research on prosodic strengthening has focused on such simple syllable structures, while little is known on prosodic strengthening in clusters (but see Bombien, Mooshammer, Hoole, Kühnert and Schneeberg 2006; Byrd and Choi 2006; Fougeron 1998). The plosives were also preceded by vowels because such a segmental context makes acoustic measurements more precise and avoids coarticulation and assimilation of the target plosive to neighboring consonants.

We focused on those cues to the fortis-lenis cues that are likely to be affected by domain-initial prosodic strengthening. Initial strengthening affects mainly consonants, whereas vowels are more affected by prominence (i.e., lexical stress and phrasal accent; Cho and Keating 2007). Moreover, initial strengthening effects are strongest in the segments immediately following the prosodic boundary and decay rapidly in subsequent segments (e.g., Bombien et al. 2006). We therefore investigated primarily the cues located in the domain-initial consonant, and took only durational measures into account for the surrounding vowels. Following the unfolding speech stream over time, we measured the duration of the preceding vowel, the closure duration, the presence of glottal vibration during the closure, the intensity of the release burst, VOT, and the duration of the following vowel. The duration of the following vowel, even though not a well established cue to the fortis-lenis contrast in German, was included because in English, vowels following word-initial fortis plosives tend to be shorter than those following lenis ones (Allen and Miller 1999). Moreover, the duration of the following vowel as the second segment after the prosodic boundary may be affected by prosodic strengthening, though possibly to a lesser extent than the preceding domain-initial plosive (Bombien et al.2006; Byrd and Choi 2006; Cho 2005; Fougeron and Keating 1997).

Both the Uniform Strengthening Account and the Feature Enhancement Account predict all fortis plosives to be more fortis-like in higher prosodic domains, that is, to be produced with shorter preceding vowels, longer closure durations with less glottal vibration, stronger releases, longer VOT, and shorter following vowels. The Uniform Strengthening Account makes the same prediction for the lenis plosives, whereas the Feature Enhancement Account predicts lenis plosives in higher prosodic domains to be more lenis like. That would imply longer preceding vowels, shorter closures with more glottal vibration, weaker releases, shorter VOT, and longer following vowels.

Method

We recorded ten native speakers of German who read sets of meaningful sentences. Embedded in these sentences were the target words, members of three minimal pairs in which the plosives /b/ versus /p/, /d/ versus /t/, and /g/ versus /k/ occurred in word-initial position. The syntactic structure of the sentences was varied in order to induce different-sized prosodic boundaries before the target words. The recordings were prosodically analyzed; thereafter we examined the effect of prosodic boundary strength on the above mentioned acoustic correlates of the fortis-lenis contrast.

Speech materials

Six German words were selected where one of the plosives /b, d, g, p, t, k/ occurred in word-initial position and was followed by /a/. These words formed three minimal word pairs, of which the members differed only in the fortis/lenis specification of their initial plosives:

(1) backen [ˈbakən] ‘to bake’ vs. packen [ˈpakən] ‘to pack’,

(2) Dank [ˈdaŋk] ‘thanks’ vs. Tank [ˈtaŋk] ‘tank’,

(3) Garten [ˈgaːtən] ‘garden’ vs. Karten [ˈkaːtən] ‘cards’.

The words that begin with /b/and /p/ are verb forms. The words starting with /d, t/ and /g, k/ were incorporated into nominal compounds, forming their first, modifying parts: *Dankschreiben* ‘letters of thanks’, *Tankdeckel* ‘tank caps’, *Gartenarbeit* ‘gardening’, and *Kartenspiele* ‘card games’. All these target words are polysyllabic and stressed on the first syllable.

The target words were embedded in meaningful sentences with four different syntactic structures which will be described in the following (see Table 1). The preceding segmental context was always the diphthong /iɐ/ in the pronoun *wir* [viɐ] 'we'. To prevent a confound of domain-initial strengthening with effects of phrasal accentuation, we induced deaccentuation of the target words by asking speakers to place a contrastive pitch accent on another (non-target) word in the utterance, as indicated in bold in Table 1.

The four sentence types were expected to be produced with various prosodic boundaries before the target words. Note that we do not assume any direct mapping between syntax and higher-level prosodic structure. The actual prosodic realization of each sentence token was classified post-hoc based on a prosodic analysis (see below). In sentences of Type A, the target words belong to the predicate group (headed by the preceding inflected verb form *wollen* 'want') of which the pronoun *wir* 'we' forms the grammatical subject. It is unlikely that speakers will produce a prosodic phrase boundary before the target words in sentences of this type and we expect only a prosodic word boundary. In sentences of Types B and C, the syntactic cohesion of the target word with the pronoun is weaker, since the target word heads a complement clause (Type B) or functions as an 'item' in a 'list' of non-finite predicate parts (Type C). This weaker cohesion is also marked in orthography by a comma. In these sentences, we can expect prosodic boundaries larger than a word boundary before the target word, but smaller than, for instance, an Intonation Phrase boundary. We chose two different sentence types to elicit these types of boundaries (Types B and C), because it is unclear from the literature (e.g., Grabe 1998; Féry 1993; Fox 1993) whether they indeed elicit intermediate prosodic boundaries (but see Kuzla, Cho and Ernestus 2007). Finally, a large prosodic boundary can be expected between the syntactically complete sentences in Type D, where orthography requires a period.

/b/	A	*Am Samstag wollen wir backen und* ***einkaufen****.* 'On Saturday, we want to do baking and **shopping**.'
	B	*Geplant hatten wir, Backen und* ***Einkauf*** *zuerst zu machen.* 'Our plan was to do baking and **shopping** first.'
	C	*Einkaufen müssen wir, backen für* ***morgen****, und aufräumen.* 'We have to go shopping, bake for **tomorrow,** and tidy up.'
	D	*Heute segeln wir. Backen kann* ***Anna****.* 'Today we go sailing. (The) Baking can be done by **Anna**.'
/p/	A	*Am Samstag wollen wir packen und* ***einkaufen****.* 'On Saturday, we want to do packing and **shopping**.'
	B	*Geplant hatten wir, Packen und* ***Einkauf*** *zuerst zu machen.* 'Our plan was to do packing and **shopping** first.'

	C	*Einkaufen müssen wir, packen für **morgen**, und aufräumen.* 'We have to go shopping, pack for **tomorrow,** and tidy up.'
	D	*Heute segeln wir. Packen kann **Anna**.* 'Today we go sailing. (The) Packing can be done by **Anna**.'
/d/	A	Am Freitag können wir Dankschreiben **drucken** lassen. 'On Friday we can have letters of thanks printed.'
	B	Geplant hatten wir, Dankschreiben **drucken** zu lassen. 'Our plan was to have letters of thanks printed.'
	C	Einkaufen müssen wir, Dankschreiben **drucken** lassen**,** und aufräumen. 'We have to go shopping, have letters of thanks printed, and tidy up.'
	D	Heute segeln wir. Dankschreiben kann **Anna** drucken lassen. 'Today we go sailing. Letters of thanks can Anna have printed.'
/t/	A	Am Freitag müssen wir Tankdeckel **zuerst** besorgen. 'On Friday we have to get tank caps first.'
	B	Geplant hatten wir, Tankdeckel **zuerst** zu besorgen. 'Our plan was to get tank caps first.'
	C	Einkaufen müssen wir, Tankdeckel **bestellen**, und aufräumen. 'We have to go shopping, order tank caps, and tidy up.'
	D	Heute segeln wir. Tankdeckel kann **Anna** besorgen**.** 'Today we go sailing. Tank caps can Anna get.'
/g/	A	Am Samstag können wir Gartenarbeit **zusammen** machen. 'On Saturday we can do (the) gardening together.'
	B	Geplant hatten wir, Gartenarbeit **zusammen** zu machen. 'Our plan was to do (the) gardening together.'
	C	Einkaufen müssen wir, Gartenarbeit **alleine** machen, und aufräumen. 'We have to go shopping, do (the) gardening alone, and tidy up.
	D	Heute segeln wir. Gartenarbeit kann **Anna** machen. 'Today we go sailing. (The) Gardening can Anna do.
/k/	A	Am Freitag wollen wir Kartenspiele **ohne** Anna machen. 'On Friday we want to play card games without Anna.'
	B	Geplant hatten wir, Kartenspiele **alleine** zu machen. 'Our plan was to play card games alone.'
	C	Einkaufen wollen wir, Kartenspiele **alleine** machen, und aufräumen. 'We have to go shopping play card games alone, and tidy up.'
	D	Heute segeln wir. Kartenspiele kann **Anna** mit euch machen. 'Today we go sailing. Card games can Anna play with you.

Table 1: Speech materials: Target words 'backen', 'packen', 'Dankschreiben', 'Tankdeckel', 'Gartenarbeit' und 'Kartenspiele' in four sentence types. Pitch-accented (non-target) words are indicated in bold.

Participants

Ten native speakers of northern Standard German, five females and five males, participated in the experiment. All of them had spent their childhood in one of the northern federal states of Germany (i.e., Schleswig-Holstein, Niedersachsen, Mecklenburg-Vorpommern, and Brandenburg), and no participant displayed any other regional accent or any speech disorder during a short interview and a reading task. They were all university students, with a mean age of 24 years, and were paid for their participation. Ülzmann (2007) reports local variations within this speaker group, with tendencies towards neutralization of the /t/-/d/ contrast in word-medial intervocalic position for speakers of Schleswig-Holstein origin. However, neutralization is unlikely to occur in our materials, since we focus on word-initial plosives and moreover, neutralization appears to be inhibited by existing lexical contrasts.

Recording procedure

The participants first read the materials silently from paper printouts. They were then recorded while reading the sentences aloud at their normal speech rate, in a fluent and natural way. They were instructed to emphasize words printed in bold, but did not receive any specific instruction on prosodic phrasing. Each sentence was repeated five times by each speaker, in randomized blocks, each consisting of one repetition of the four sentences per plosive. In total, 1200 sentence tokens were recorded. Recordings were made in a sound-attenuated booth with a Sennheiser MD 421 microphone and digitized directly into a computer at a sampling rate of 16 kHz.

Prosodic categorization

Given that there is no one-to-one correspondence between syntactic and prosodic structure, we based the classification of prosodic boundaries entirely on temporal and intonational criteria. Research on the perceptual strength of prosodic boundaries (de Pijper and Sanderman 1994) has shown that the presence of a pause is sufficient to signal a high prosodic boundary. Another strong indicator of prosodic boundaries above the word level is a non-accent-lending pitch movement or 'boundary tone'. Prosodic word boundaries typically lack such melodic markers. Combining these criteria, we defined three prosodic categories to which we assigned our data: the *Major* category was characterized by the presence of a pause and a boundary tone. A pause was defined as a period of silence of at least 150 ms. In addition, 21 tokens were coded as containing a pause based on the perceptual judgments of the two independent transcribers. The *Minor* category was characterized by a boundary tone, but no pause, and the *Word* category by the absence of both a pause and a boundary tone.

These criteria have been applied in other studies before (Bombien et al. 2006; Cho and McQueen 2005; Kuzla et al. 2007; Kuzla, Ernestus and Mitterer in press), and the resulting boundaries correspond roughly to the intonation phrase, the intermediate phrase, and the prosodic word boundaries of Beckman and Pierrehumbert (1986). Two trained native listeners coded the boundary tones separately, with 92.7 % agreement. Three tokens that remained ambiguous after re-inspection were excluded from analysis, as were 19 accented target words. The final data set consisted of 1178 tokens.

The four sentence types induced indeed the expected variation in prosodic realizations, with considerable differences between speakers in their preferred phrasing strategies in particular for sentence types B and C. In addition, there was also within-speaker variation. The distribution of sentence types over prosodic categories is given in Table 2. In general, sentence type A elicited Word boundaries and sentence type D Major boundaries. Types B and C showed more variation, but there was a general preference for Minor boundaries in Type B and a slight preference for Major boundaries in Type C.

	Major	*Minor*	*Word*
Type A : Word	0	2	290
Type B: Complement clause	97	181	20
Type C : List	160	137	2
Type D: Sentence	228	61	0

Table 2: Distribution of sentence types over prosodic categories.

Acoustic measurements

A phonetically trained labeler annotated the waveform and spectrogram of the recordings in PRAAT (Boersma 2001), determining acoustic landmarks in the target word's onset syllable and in the preceding syllable. First, we measured the duration of the syllable *wir* /viɐ /, which preceded the prosodic boundary and was often realized as a labial approximant followed by a centralized mid-open vowel. Second, we determined the duration of the stop closure. Note that in the presence of a preceding pause, the beginning of the closure cannot be determined, and accordingly, we did not study closures preceded by a Major prosodic boundary. Third, we measured the duration of glottal vibration during the closure and calculated the percentage of the closure duration produced with glottal vibration for each token. Fourth, we measured the VOT, defined as the interval from the closure release to the onset of high-amplitude periodicity in the waveform for the following vowel. Fifth, for the fortis plosives, we also extracted the intensity maximum in the first 15 ms of the release noise (for the lenis plosives, we could not obtain enough reliable measurements because many tokens were produced with weak releases shorter than 15 ms). Finally, we measured the duration of the following vowel.

Results

For all analyses, we built multi level regression models (e.g., Bates 2005; Baayen 2008; Dalgaard 2002) with Prosodic Category (henceforth: PCat; 3 levels: Major, Minor, Word), Place of Articulation (Place; 3 levels: labial, alveolar, velar), and the fortis-lenis contrast (Contrast; 2 levels: fortis, lenis) as fixed factors, and with Speaker as a random variable. This type of statistical models acknowledges the insight that speakers differ considerably in their articulation, including for instance speech rate, amount of glottal vibration, and loudness. While estimating the effects of the fixed predictors on the dependent variable, a different intercept for each participant is calculated. Thus, these models reduce the variance in the data contributed by inter-speaker differences, and are able to detect effects of the fixed factors which might not be captured by a comparison of group means only.

Effects of Prosodic Boundary

Preboundary lengthening

We analyzed the preboundary syllable duration for two reasons. First, such an analysis allows us to evaluate our prosodic classification, since preboundary lengthening is a well-established correlate of prosodic structure (e.g., Wightman, Shattuck-Hufnagel, Ostendorf and Price 1992). If our prosodic categories appear to be good predictors of the duration of the preboundary syllable *wir*, we take that as external evidence for their validity (the same procedure has been applied in Kuzla et al. 2007).

Second, at the same time we can test for an effect of the fortis/lenis contrast on the duration of the preceding vowel. Within words, the duration of the preceding vowel is a cue to the fortis/lenis contrast of the following consonant in German (Kohler 1977, 1979; Piroth and Janker 2004), since (given a constant speech rate, as in our experiment) vowels tend to be shorter before fortis than before lenis consonants. We may find no effect of the phonological contrast in our data since the vowel and the plosive do not belong to the same prosodic word. Analysis of the preboundary syllable duration as a function of PCat, Contrast and Place yielded significant main effects of PCat ($F(2, 1160) = 1122.91$; $p < 0.001$) and Place ($F(2, 1160) = 3.19$; $p < 0.05$), but no effect of Contrast and no interactions. Additional analyses showed that the effect of Place was due to /g/, for which the preboundary syllable was slightly longer (on average 9.65 ms) than for the other plosives. Importantly, all prosodic categories differed from each other in the expected direction (all $p < 0.001$; means: Major 230 ms, Minor 118 ms, Word 99 ms). We take this result as evidence for the validity of our prosodic categorization. That there is no effect of Contrast on preboundary lengthening indicates that in sequences of a vowel and a plosive separated by a prosodic word boundary (V#C), vowel duration is not a cue to the fortis-lenis contrast, as opposed to the VC# case.

Closure duration

We analyzed the duration of the plosive closure for the prosodic categories Minor and Word only, since the beginning of the closure could not be determined after the pause in the Major condition. PCat ($F(1,681) = 508.34$; $p < 0.001$) and Place ($F(2,681) = 73.55$; $p < 0.001$) emerged as significant, as did the interaction between PCat and Contrast ($F(1,681) = 4.04$; $p < 0.05$). To investigate the interaction, we split the data by Contrast.

For fortis plosives, PCat and Place were significant predictors (PCat: $F(1,352) = 211.64$; $p < 0.001$; Place: $F(2,352) = 39.85$; $p < 0.001$). The average closure durations by prosodic category and place of articulation are given in Table 3. As expected, closures were longer at the higher prosodic boundary (Minor > Word). Furthermore, pair-wise comparisons showed that closure duration differed significantly (all $p < 0.001$, Bonferroni corrected) between the three places of articulation (/p/ > /t/ > /k/).

	/p/	/t/	/k/
Minor	107 (20.9)	100 (20.4)	88 (28.0)
Word	80 (15.8)	69 (11.7)	56 (13.7)

Table 3: Mean closure duration [ms] of fortis plosives as a function of prosodic category and place of articulation. Standard deviations are given in brackets.

The analysis of the lenis plosives also yielded a main effect of Place ($F(2,329) = 32.97$; $p < 0.001$). The three places of articulation differed from each other in the same direction as for the fortis plosives (all $p < 0.001$; /b/ > /d/ > /g/). Additionally, we observed a main effect of PCat ($F(1, 329) = 304.79$; $p < 0.001$), with again longer closures in the Minor condition. Means are provided in Table 4. As illustrated in Figure 1, this effect of prosodic category is slightly larger for the lenis than for the fortis plosives, which explains the interaction in the main analysis.

	/b/	/d/	/g/
Minor	114 (22.2)	100 (28.5)	94 (26.2)
Word	77 (13.3)	63 (13.2)	55 (12.3)

Table 4: Mean closure duration [ms] of lenis plosives as a function of prosodic category and place of articulation. Standard deviations are given in brackets.

These findings do not allow us to evaluate the Uniform Strengthening Account and the Feature Enhancement Account, since closure duration does not appear to be a cue to the fortis-lenis contrast of word-initial plosives. The analysis showed no main effect of Contrast, and also Jessen (1998) found no effect of closure duration for word-initial plosives, in contrast to word-medial plosives. We will come back to this issue below.

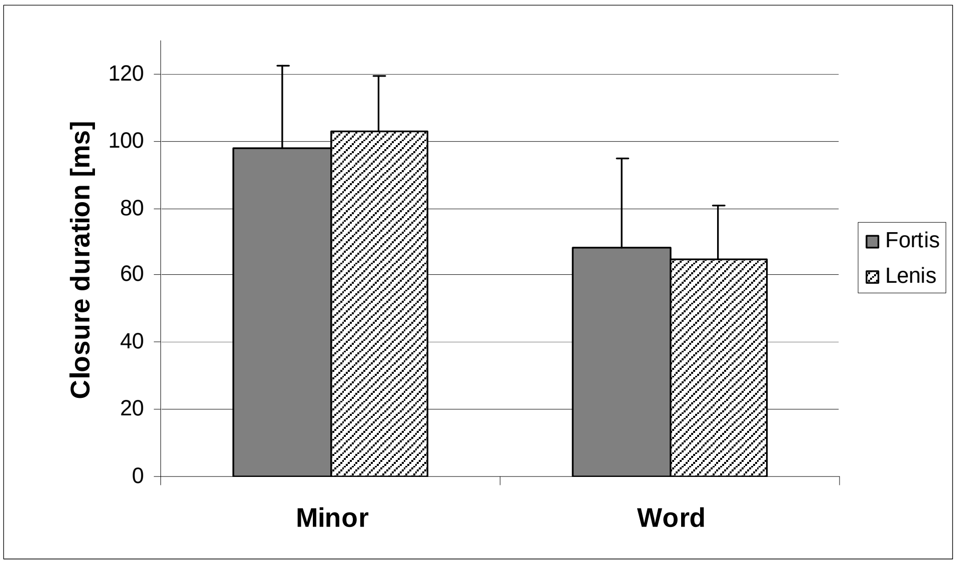


Figure 1: Closure duration as a function of Prosodic Category and Contrast.

Voice Onset Time

For Voice Onset Time (Figure 2), there were main effects of PCat ($F(2,1160) = 34.45$; $p < 0.001$), Place ($F(2,1160) = 139.32$; $p < 0.001$), and Contrast ($F(1,1160) = 4084$; $p < 0.001$), and a significant interaction of PCat and Contrast ($F(2,1160) = 25.40$; $p < 0.001$). To investigate this interaction, we split the data again by Contrast and modeled VOT as a function of PCat and Place.

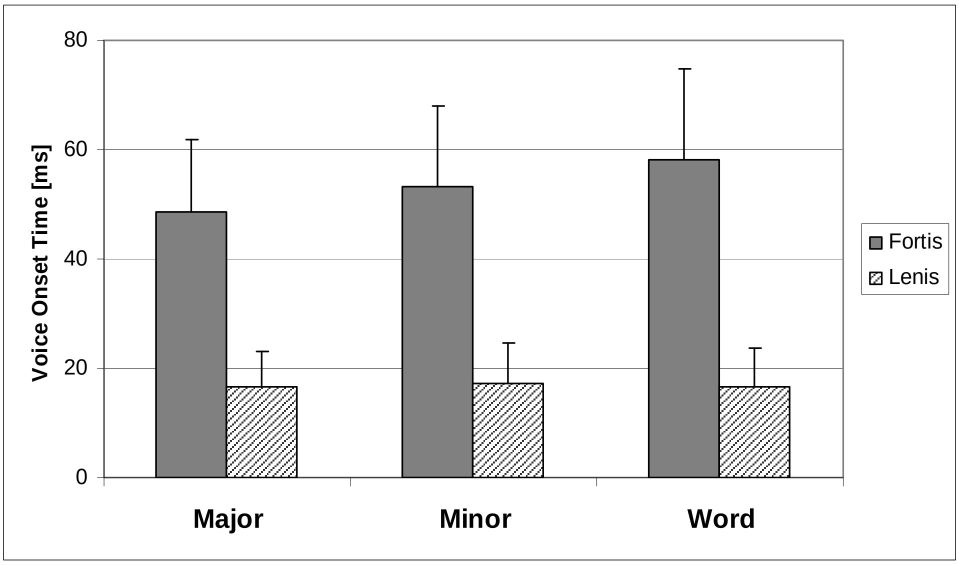


Figure 2: Voice Onset Time as a function of Prosodic Category and Contrast.

For the fortis plosives, there were significant main effects of PCat ($F(2,583) = 28.59$; $p < 0.001$) and Place ($F(2,583) = 57.78$; $p < 0.001$). Across all places of articulation, VOT was longest after Word boundaries (mean: 58.1 ms), shorter after Minor boundaries (mean: 53.4 ms), and shortest after Major boundaries (mean: 48.5 ms), with all differences between prosodic conditions being significant (all $p < 0.001$). Furthermore, pair-wise comparisons showed that VOT was different for the three places of articulation (/p/ < /t/ < /k/; all $p < 0.001$), mean values being 47 ms for /p/, 51 ms for /t/, and 60 ms for /k/. This order is as expected, given that a larger oral cavity facilitates vocal fold vibration, as it allows for a lower supraglottal pressure during the stop closure phase.

For the lenis plosives, there was no effect of PCat ($F(2, 585) = 1.19$; $p > 0.1$), which explains the interaction in the overall analysis. Only the effect of Place was significant ($F(2, 585) = 227.38$; $p < 0.001$). Pair-wise comparisons revealed that all places of articulation differed from each other (all $p < 0.001$), showing a pattern of /b/ (mean: 12 ms) having a shorter VOT than /d/ (mean: 16 ms), both having shorter VOT than /g/ (mean: 22 ms), which is the same pattern as for the fortis plosives.

Our finding that VOT for fortis plosives decreases with stronger prosodic boundaries is exactly the opposite of the predictions made by both the Uniform Strengthening Account and the Feature Enhancement Account. The absence of a prosodic effect for lenis plosives is also unexpected under the Uniform Strengthening Assumption. We have to conclude that the two accounts do not appear to account well for the effects of domain-initial strengthening in German.

Intensity of release noise (fortis plosives)

For the fortis plosives, we analyzed the intensity maximum in dB during the first 15ms of the release noise as a function of PCat and Place. Both factors and the interaction emerged as significant (PCat: $F(2, 579) = 4.23$; $p < 0.05$; Place: $F(2,579) =225.77$; $p < 0.001$, PCat : Place: $F(4,579) = 3.72$; $p < 0.01$). Mean intensity maximum values by prosodic category and place of articulation are provided in Table 5.

	/p/	/t/	/k/
Major	55.1 (4.4)	58.4 (3.3)	59.6 (3.6)
Minor	53.7 (3.4)	58.4 (3.4)	60.6 (5.0)
Word	54.2 (4.0)	59.2 (3.3)	61.5 (4.2)

Table 5: Mean intensity maximum [dB] of release noise in fortis plosives as a function of prosodic category and place of articulation. Standard deviations are given in brackets.

Because of the interaction, we analyzed the effect of PCat separately for each plosive. The effect of prosodic category was absent for /p/ ($F(2,193) = 0.72$; $p > 0.1$), but present for /t/ ($F(2,190) = 3.21$, $p < 0.05$) and /k/ ($F(2,196) = 9.82$; $p < 0.001$). For both /t/ and /k/, the release was slightly, but significantly softer (mean difference: 1.4 and 0.8 dB respectively) in the Major condition than in the Word condition (both $p < 0.01$). For /k/ the difference between the Major and the Minor condition was significant as well ($p < 0.001$, mean difference: 1.0 dB). Possibly, we found no effect for /p/, because VOT was shorter than 30 ms for no less than 10% of these tokens, which may have made the estimation of the maximum intensity less reliable.

Both the Uniform Strengthening Account and the Feature Enhancement Account predict stronger release noises at higher prosodic boundaries, that is, exactly the opposite of what we found for /t/ and /k/. These data appear to support our conclusion based on the analysis of VOT (section 3.1.3.) that the two accounts cannot capture well the effect of domain-initial strengthening on plosive releases in German.

Glottal vibration

We analyzed glottal vibration during the closures of the plosives. Since we could not measure the closures in the Major boundary condition, we restricted our analyses to the Minor and Word conditions only. First, we examined to what extent glottal vibration occurred during the closure, as expressed in the percentage (%) of the closure duration produced with glottal vibration. We analyzed the percentage of voicing as a function of PCat, Place of articulation, and Contrast, with Speaker as a random variable. All predictors showed main effects (PCat: $F(1,686) = 199.05$; $p < 0.001$, Place: $F(2,686) =3.84$; $p < 0.05$, Contrast: $F(1,686) = 228.37$; p

< 0.001), and also the interaction of PCat with Contrast was significant ($F(1,686) = 33.42$; $p < 0.001$). Alveolar plosives were produced with higher percentages of glottal vibration than bilabial and velar plosives (all $p < 0.05$), which did not differ from each other ($p > 0.1$). This somewhat unexpected result, given aerodynamics, is probably due to the longer duration of bilabials compared to alveolars and velars, which results in smaller percentages of glottal vibration given the same absolute duration of glottal vibration. We will come back to this in the following.

To investigate the role of prosody for the fortis and the lenis plosives, we split up the data set by Contrast. The fortis plosives showed main effects of PCat ($F(1,354) = 57.43$; $p < 0.001$) and of Place ($F(2,354) = 4.15$; $p < 0.05$). The fortis plosives following Word boundaries showed a higher percentage of glottal vibration than those following Minor boundaries (means: 19% for Minor and 30% for Word). Close inspection of the data, however, revealed that this was completely due to differences in closure duration between the two conditions. Plosives in the two conditions showed the same absolute amount of glottal vibration, but because of the longer closure duration, this resulted in lower percentage values in the Minor condition. The Place effect revealed the same pattern as the analysis of the entire dataset, alveolars being produced with a larger percentage of glottal vibration (on average 37%) than both bilabials and velars, which did not differ from each other (both 33% on average).

For the lenis plosives, there was no effect of Place, but PCat emerged again as significant ($F(1,332) = 144.4$, $p < 0.001$), with a mean of 32% of the closure produced with glottal vibration for Minor and 60% for Word (Figure 3). Under the assumption that the presence of glottal vibration is a sufficient (though not a necessary) cue to the fortis-lenis distinction, this direction of the effect of PCat is unexpected under a Feature Enhancement Account. This account would predict that lenis plosives become more lenis and therefore that the percentage of glottal vibration increases with domain-initial strengthening. In contrast, the result is in line with the predictions of the Uniform Strengthening Account.

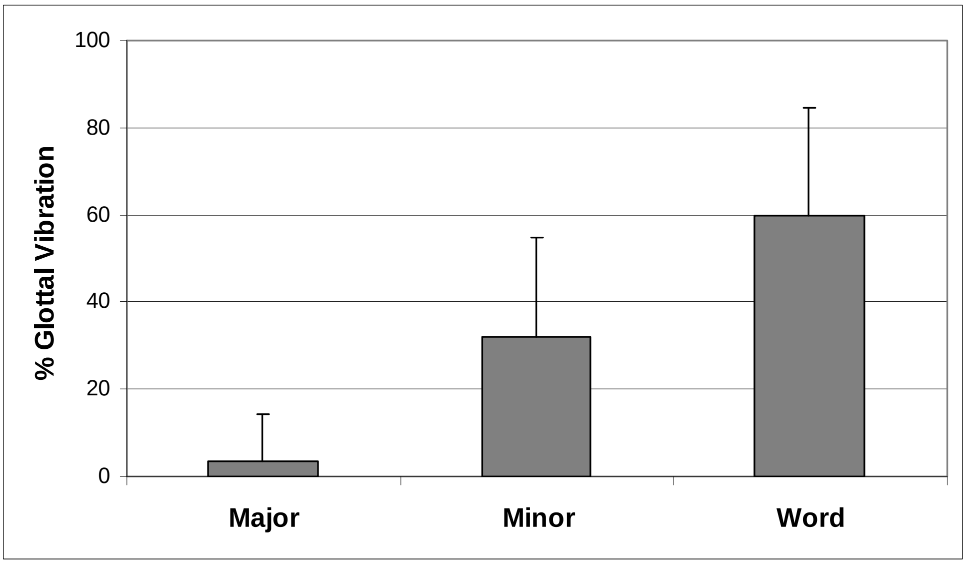


Figure 3: Percentage of glottal vibration during lenis plosives as a function of Prosodic Category.

Obviously, shorter closure durations favor higher percentages of glottal vibration: the same absolute duration of glottal vibration results in a higher percentage of glottal vibration for shorter plosives. Hence, it is possible that it is actually the closure duration which is driving the observed effect of PCat on glottal vibration (as it was the case for the fortis plosives). To test for this possibility, we performed an additional analysis. We built two models in which both PCat and Closure duration were predictors of the percentage of glottal vibration. The two models only differ in the order of the two predictors. PCat emerged as significant in the first model, where it was the first predictor ($F(1,300) = 125.71$; $p < 0.001$), and so did Closure duration as the second predictor ($F(1,300) = 73.67$; $p < 0.001$). Both predictors appeared also significant in the second model, where Closure duration was the first predictor ($F(1,300) = 189.29$; $p < 0.001$), and PCat was the second predictor ($F(1, 300) = 10.08$; $p < 0.01$). These results show that the effect of PCat on the amount of glottal vibration is robust and cannot be entirely ascribed to longer closure durations at higher boundaries.

Recent studies (e.g., Jaeger in press) suggest that percentages do not form a reliable dependent variable in statistical analyses. Moreover, also our own analysis of the fortis plosives suggests that the analyses of absolute values of glottal vibration with closure duration as a covariate provide better insight into the effects of Place and PCat on glottal vibration. In addition to the above reported analysis, which is in line with previous analyses of glottal vibration (e.g., Kuzla et al. 2007), we therefore also performed an analysis of the absolute duration of glottal vibration (in ms) on the same dataset of lenis plosives (see

Ernestus, Lahey, Verhees and Baayen 2006 for a similar analysis). We built a multi-level regression model of the duration of glottal vibration as a function of PCat and Place. As a third predictor, we entered Closure duration. For aerodynamic reasons, we expected a nonlinear relationship of glottal vibration and closure duration: With increasing closure duration, it will be harder to maintain glottal vibration. Therefore, we also added the square of the closure duration as a predictor. Finally, in order to obtain less skewed distributions of closure duration and of glottal vibration, these variables were log-transformed. The results are given in Table 6.

Fixed effects:

	Estimate	Std.Error	t-value
Intercept	-36.2204	10.7055	-3.383
logdur	18.101	4.627	3.912
(logdur^2)	-2.0591	0.5013	-4.107
PCat_Word	6.3171	2.5765	2.452
Place_alveolar	0.4274	1.9308	0.221
Place_velar	4.2515	2.0487	2.075
logdur:PCat_Word	-1.4266	0.6	-2.377
logdur:Place_alveolar	-0.1544	0.4286	-0.36
logdur:Place_velar	-1.1216	0.4564	-2.457

Table 6: Results of mixed-model analysis of the absolute voicing duration [ms] in lenis plosives as a function of the log-transformed closure duration (logdur), squared log-transformed closure duration (logdur^2), Prosodic Category and Place of articulation. The Intercept is for bilabial plosives following Minor boundaries (contrast coding).

As expected, there is a nonlinear relationship between the duration of glottal vibration and of the closure duration (simple term: $F(1, 325) = 7.99$, $p < 0.01$; squared term: $F(1, 325) = 33.71$, $p < 0.001$). Importantly, there is also a main effect of PCat ($F(1,325) = 7.00$; $p < 0.05$) and an interaction of PCat with Closure duration ($F(1,325) = 5.13$; $p < 0.05$), which is illustrated in Figure 4. Plosives with relatively short to medium closure duration (up to about 80 ms) were produced with longer glottal vibration after a Word boundary (replicating the results of the analysis of percentage of voicing reported above). For relatively long plosives, the graph suggests the opposite, which is, longer glottal vibration after a Minor boundary. However, additional analyses yielded no effect of PCat for plosives with closures longer than 80 ms. The seeming difference between Minor and Word boundaries in the right part of Figure 4 appears thus to result only from the quadratic function in the model.

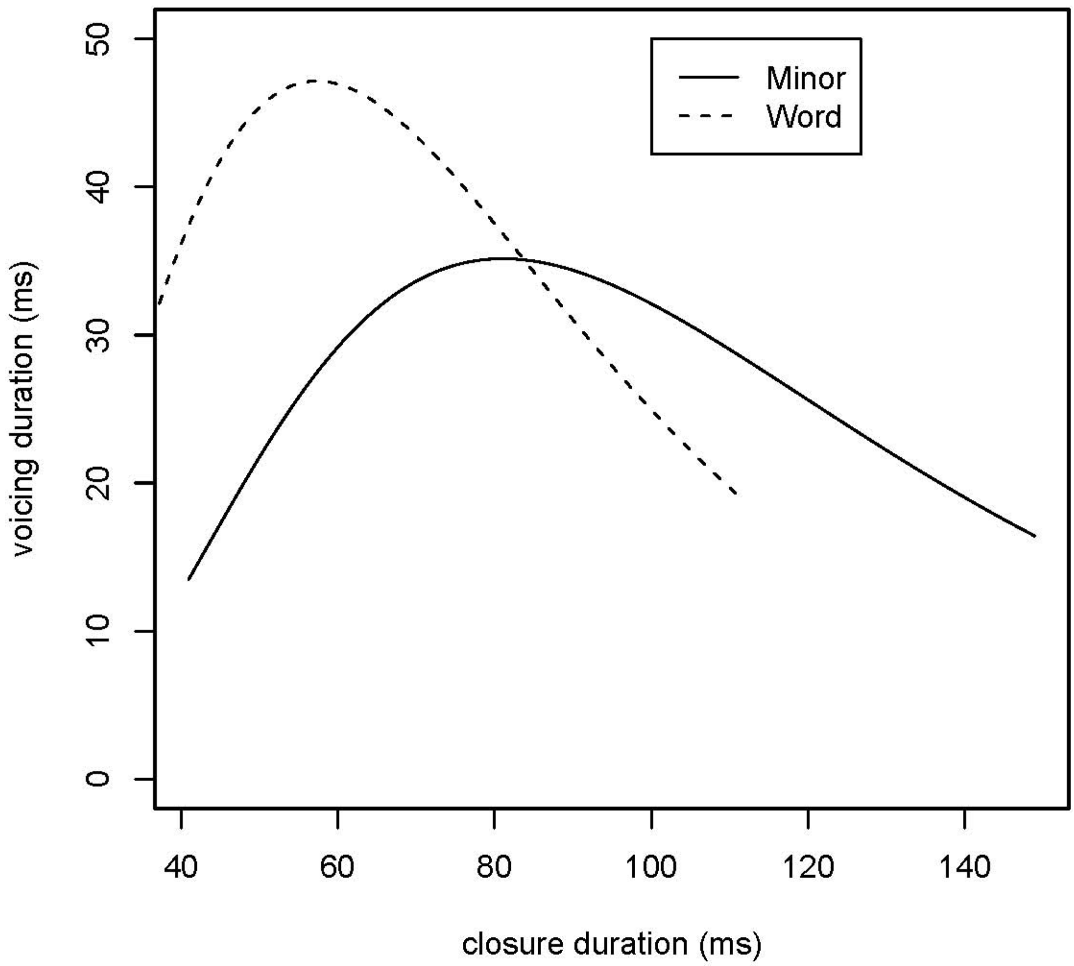


Figure 4: Correlation of absolute voicing duration and closure duration in lenis plosives for the prosodic conditions Minor and Word. (Place of articulation is bilabial. Mainly the intercept changes for alveolar and velar plosives, see Table 6.)

In the analysis, there was also an effect of place of articulation ($F(2,325) = 11.37$; $p < 0.001$), in addition to an interaction of Place with Closure duration ($F(2,325) = 4.67$, $p < 0.05$). As expected on the basis of aerodynamic differences due to oral cavity size and supraglottal pressure, there was longer glottal vibration in bilabial and alveolar plosives than in velar plosives. These results of Place of articulation clearly show the relevance to analyze the absolute duration of glottal vibration with closure duration as a covariate.

In conclusion, plosives following Minor versus Word boundaries differ in their absolute amount of voicing especially if they are short. The difference disappears for longer closures, probably for aerodynamic reasons. These data, showing more voicing after smaller boundaries, are therefore still hard to reconcile with the Feature Enhancement Account.

Duration of the following vowel

For all plosives, we analyzed the duration of the following vowel as a function of PCat, Contrast and Place. All three factors emerged as significant (PCat: $F(2,1165) = 4.43$; $p < 0.05$, Contrast: $F(1,1165) = 331.80$; $p < 0.001$, Place: $F(2,1165) = 2117.26$; $p < 0.001$), as did the interactions between PCat and Place ($F(4, 1165) = 5.05$; $p < 0.001$) and Contrast and Place ($F(2,1165) = 43.27$; $p < 0.001$). To investigate these interactions, we modeled the duration of the following vowel as a function of PCat and Contrast for each place of articulation separately. Contrast was significant in all analyses (all $p < 0.001$): Vowels following fortis plosives were on average 14 ms shorter than vowels following lenis plosives (means: 76 ms versus 90 ms). The effect of PCat was present for velars only ($F(2,394) = 7.60$; $p < 0.001$). For velars, vowels were significantly longer after prosodic word boundaries (all $p < 0.001$ in pair-wise comparisons) than after both Minor and Major phrase boundaries, which did not differ from each other ($p > 0.1$). The mean durations of the following vowel after a velar were 120 ms for Word boundaries, 115 ms for Minor and 113 ms for Major boundaries.

That we observed an effect in the vowels following velars, but not in those following alveolars and labials, may be attributed to the fact that in our materials, vowels following velars had overall longer durations. The vowels in *Garten*- and *Karten*- are tense and followed by /r/, which is vocalized in German in postvocalic position, resulting in an open syllable structure, whereas the vowels in *backen, packen, Dank*- und *Tank*- are lax and occur in closed syllables. These latter vowels may have been too short for the prosodic boundary effect to emerge.

The attested prosodic effect for vowels following velars is compatible with the Uniform Strengthening Account, which predicts that vowels following both fortis and lenis plosives get shorter at higher boundaries. The result is not in line with the Feature Enhancement Account, as it predicts that domain-initial strengthening leads to shorter vowels after fortis plosives but to longer vowels after lenis plosives. However, as the prosodic boundary effect is limited to vowels after velars in our data, it does not permit strong conclusions with respect to either account.

Predictors of Contrast

The above discussions of the different acoustic cues to the fortis-lenis distinction suggest that neither the Uniform Strengthening Account nor the Feature Enhancement Account can completely explain our data. Possibly prosodic structure does not affect the acoustic cues one by one independently in the direction predicted by the Uniform Strengthening or the Feature Enhancement accounts but takes into account the relevance of the different cues for the fortis-lenis contrast. In order to evaluate this possibility we need to know the relevance of these cues.

We investigated which of the examined acoustic characteristics do indeed cue the fortis-lenis distinction, and what their relative contributions are. We built a generalized multi-level model, predicting the fortis-lenis specification of the plosives (as intended by the speakers according to the presented stimuli) as a function of the duration of the preceding vowel, closure duration, VOT, the log of the percentage of the closure produced with glottal vibration plus 1 (we took the log in order to obtain an approximately normal distribution of this variable, and we added 1 to avoid taking the log of zero for completely voiceless plosives), and the duration of the following vowel, again with Speaker as a random factor. We restricted the data set to the plosives following only Word or Minor boundaries, since closure duration data are not available after pauses in the Major condition. Furthermore, we excluded as fixed predictor the intensity maximum of the burst, as no measurements were obtained for the lenis plosives. Note that the fact that these measurements were successfully carried out for the fortis plosives, but not for the lenis plosives, suggests that burst intensity is a valuable cue to the fortis/lenis distinction.

We expected no contribution from the duration of the preceding vowel, as its duration is mainly determined by domain-final lengthening (see above). The role of closure duration in the fortis-lenis contrast is uncertain. Previous studies have shown that it is an important cue for word-medial plosives (Fischer-Jørgensen 1976; Jessen 1998). However, its contribution to word-initial, but utterance-medial plosive contrasts may be reduced, as is also suggested by our data on closure duration, which showed no difference in duration between the fortis and lenis plosives.

Voicing is not a necessary, but if present, a sufficient cue to lenis plosives (Jessen 1998; 2001). We therefore expect an effect of glottal vibration. We also expect an effect of VOT, since it has been shown to be a primary correlate of the fortis-lenis distinction (Fuchs 2005; Jessen 1998, Ülzmann 2007), and because it is longer for fortis than for lenis plosives also in our data (see section 3.1.3.). Whereas the duration of the preceding vowel is probably irrelevant, the duration of the following vowel may play a role, as is also suggested by our data reported in the respective section above.

The analysis showed that there were only two significant predictors: VOT ($F(1,689) = 87.323$, $p < 0.001$) and the duration of the following vowel ($F(1,689) = 28.75$, $p < 0.001$). We found no effect of the duration of the preceding vowel, of closure duration, or of the percentage of glottal vibration (all $p > 0.1$). Hence, even though we observed an interaction of Contrast with PCat on closure duration in the analysis reported above, indicating that the PCat effect was larger for lenis than for fortis plosives, closure duration by itself appears no strong cue to the contrast.

The absence of an effect of the percentage of glottal vibration is more surprising. The results are very similar if percentage of voicing is replaced by the mere presence versus absence of voicing, or by the absolute duration of glottal vibration. None of these alternative predictors reached significance in according analyses, which we therefore do not report here in detail. Possibly, the absence of an effect of glottal vibration is due to the large number of plosives with 0% or 100% voicing in the data set, which complicates the testing of this predictor. Indeed, this predictor is significant ($F(1,574) =17.14$, $p < 0.001$) if the tokens with 0% or 100% voicing are removed from the data set.

We then investigated the source of the effect of the following vowel duration. This effect may result from a complementary division of the time span from the release of the plosive to the end of the vowel between aspiration and vowel. The duration of the following vowel is then negatively related with VOT, and its contribution to the prediction of the contrast may be completely due to this correlation. Indeed there was a significant correlation between the duration of the following vowel and VOT (Pearson correlation, $r = -0.15$, $p < 0.001$). In order to investigate whether the duration of the following vowel also contributed to the predictability of the contrast independently, we created a new variable that is highly correlated with the duration of the following vowel, but which is not correlated with VOT. We modeled the duration of the following vowel as a function of VOT and called the residuals of this model the duration residuals. The duration residuals show a Pearson correlation of 0.97 with the duration of the following vowel ($p < 0.001$), but a non-significant correlation of almost 0.00001 with VOT. We then modeled the Contrast as a function of VOT and the duration residuals, with Speaker as a random variable. VOT was still a significant predictor ($F(1, 689) = 85.09$, $p < 0.001$). More importantly, also the duration residuals emerged as significant ($F(1, 689) = 28.23$, $p < 0.001$), which shows that the duration of the following vowel contributes to the voicing contrast, independently from VOT.

Not surprisingly, VOT had a much greater effect. The maximal effect of VOT (i.e., the maximal value of VOT multiplied by the beta coefficient, which is the estimated effect of the predictor, see e.g., Chatterjee, Hadi and Price 2000) on the logit of Contrast was 45, whereas the effect of the vowel was much smaller (7). VOT appears thus to be the most important predictor of the fortis-lenis contrast.

General Discussion

The present study investigated the influence of prosodic structure on the phonetic realization of fortis and lenis plosives in German. In particular, we addressed the question how prosodic strengthening affects acoustic cues to the fortis-lenis distinction. In the previous literature, two lines of explanation have been put forward: the Uniform Strengthening Account and the Feature Enhancement Account of prosodic strengthening. While the Uniform Strengthening Account predicts that prosodic strengthening makes both fortis and lenis plosives in general more fortis-like in stronger prosodic positions, the Feature Enhancement Account predicts lenis cues to become stronger for the lenis plosives, and fortis cues to become stronger for fortis plosives.

Our data consisted of sentences read by ten speakers, with different prosodic boundaries before the word-initial plosives. We investigated the durations of the preceding vowel, of the consonant closure, of the following vowel and VOT for both the fortis and lenis plosives as well as the amount of glottal vibration in lenis plosives (in relative as well as absolute measures), and the burst intensity maximum in fortis plosives. All these measures showed main effects of prosody, as indicated in the first two columns of Table 7.

Acoustic Characteristic	***Prosodic Effect***	***Uniform Strengthening***	***Feature Enhancement***
Preceding vowel	longer at higher boundaries	inconclusive	inconclusive
Closure duration	longer at higher boundaries	inconclusive	inconclusive
VOT	shorter for /p, t, k/ at higher boundaries	contra	contra
Burst intensity	lower at higher boundaries	contra	contra
Glottal vibration	less at higher boundaries	pro	contra
Duration of following vowel	longer at smaller boundaries (for velars only)	pro, if not inconclusive	inconclusive

Table 7: Prosodic effects on potential cues to the fortis-lenis distinction and their compatibility with the Uniform Strengthening account and the Feature Enhancement account of prosodic strengthening.

Before we interpreted the results, we investigated whether all analyzed acoustic characteristics are indeed cues to the fortis/lenis distinction in German. We built a model predicting the Contrast value (fortis/lenis) of the plosive as intended by the speaker. All acoustic characteristics appeared to be correlated with this fortis/lenis value, except the duration of the preceding vowel and the duration of the closure. The duration of the preceding vowel is not correlated probably because it is separated from the plosive by at least a prosodic word boundary. The absence of a correlation for closure duration is in line with earlier findings by Jessen (1998) for word-initial plosives and suggests that closure duration is only an acoustic cue for the fortis/lenis identity of plosives in word-medial position in German. Among the acoustic characteristics that are cueing the fortis/lenis distinction, VOT appeared most important.

We will now discuss the prosodic effects on the four characteristics that are cues to the fortis/lenis distinction and their implications for the Uniform Strengthening Account and the Feature Enhancement Account (see the two right most columns of Table 7). Note that we wrote "inconclusive" in this table if the acoustic characteristic appeared not to cue the fortis/lenis distinction, or if the accounts do not make predictions for these characteristics.

First, VOT decreased at higher boundaries for fortis plosives, whereas it was not affected by prosodic boundary size for lenis plosives. The decrease in VOT for fortis plosives is unexpected under the Uniform Strengthening and the Feature Enhancement accounts. It is true that our results are similar to the findings for Dutch reported by Cho and McQueen (2005), but in German, fortis plosives are assumed to be specified by {+spread glottis}, in contrast to {-spread glottis} in Dutch. Together with the null effect of prosody on VOT in lenis plosives, this pattern implies contrast diminishing rather than contrast enhancement at higher prosodic boundaries.

Second, we observed higher burst maxima in lower prosodic domains, in contrast to the predictions by both the Uniform Strengthening and the Feature Enhancement accounts. Interestingly, a similar result has been reported for English by Cho and Keating (2007) who measured high-pass filtered RMS energy in /t/. These authors attribute the higher burst energy found at smaller boundaries to possible inverse correlations between the burst amplitude and other properties of the preceding closure which may have undergone articulatory strengthening, such as amount of articulatory contact and velocity of the tongue release. This would make burst maximum results inconclusive for the evaluation of the two prosodic strengthening accounts, as we can only speculate about the involved articulatory gestures.

Third, we found more glottal vibration in the (short to medium length) lenis plosives at smaller boundaries. Glottal vibration is not considered a necessary cue for the fortis/lenis distinction in German plosives (that is why there is no distinctive feature [voice]), but nevertheless it contributes to the perception of the fortis/lenis contrast (e.g., Jessen 1998). More glottal vibration enhances the lenis character of plosives. In contrast to the predictions of the Feature Enhancement Account, but in line with the Uniform Strengthening Account, we thus see that reduced glottal vibration makes lenis plosives less lenis like in prosodically stronger locations.

Finally, the duration of the following vowel was hardly affected by prosodic structure. Only for velars, we observed longer vowel durations after smaller prosodic boundaries. It has been observed previously (e.g., Cho 2005; Cho and Keating 2007; Fougeron and Keating 1997; Fougeron 2001) that domain-initial strengthening may be located primarily in the very initial segment after the prosodic boundary and that the effects decay rapidly in the following segments. The direction of the effect in the vowels following velars is unexpected under a simple domain-initial strengthening account, which predicts that segments are strengthened if they are close to a stronger prosodic boundary: Temporal expansion at higher prosodic boundaries would lead to longer vowel durations, which is the opposite of what we found. However, the result is in line with the Uniform Strengthening Hypothesis if we take into account that vowels following fortis plosives tend to be shorter than those following lenis plosives (see our analysis in section 3.2. and the findings by Allen and Miller 1999): Prosodically stronger boundaries lead to vowel durations which are more typical for fortis plosives.

Overall, our findings do not consistently support either of the two accounts of prosodic strengthening. Rather, we see that prosodic structure does not affect the individual cues in a uniform way. That is, some acoustic characteristics suggest that plosives (either lenis or fortis) are more fortis like after stronger prosodic boundaries (e.g., the shorter duration of the following vowel, reduced glottal vibration), whereas other cues make the fortis plosives more lenis like (lower burst intensity and shorter VOT). As a result of this, contrasts appear to be maintained in general, not only within a given prosodic position, but also across positions.

Recently, Cole et al. (2007) have found also that prosodic strengthening is generally contrast-preserving. In their data of four Radio News speakers, accent strengthening effects were greatest for those acoustic properties that play smaller roles in encoding phonological contrasts. They concluded that acoustic properties which carry a heavy functional load in terms of phonological contrasts are less free to serve pragmatic functions and are therefore also less susceptible to prosodically conditioned variation induced by phrasal accentuation. Cole et al. (2007) formulated this as a principle of 'phonology-over-pragmatics'.

This hypothesis can be extended to domain-initial strengthening. Possibly, domain-initial strengthening also affects especially those characteristics of the acoustic signal that are not highly relevant for distinguishing between phonological contrasts. We will refer to this hypothesis as the contrast-over-prosody hypothesis.

According to our results, the acoustic correlates of the fortis/lenis contrast are VOT, following vowel duration, glottal vibration, and presumably, the intensity maximum of the burst (we did not test this latter cue, but given that it could only be measured for fortis plosives, it is evident that it distinguishes between fortis and lenis plosives). The contrast-over-prosody hypothesis predicts that there are only limited prosodic effects on these correlates. On the other hand, duration of the preceding (preboundary) vowel, and closure duration appear not to contribute to the fortis-lenis distinction, and thus the contrast-over-prosody hypothesis predicts that the prosodic effects on these acoustic characteristics of the signal are greater (assuming that they do not cue other important phonological contrasts).

The two measured acoustic characteristics that do not contribute to the fortis/lenis distinction, the duration of the preceding vowel and the closure duration, showed huge effects of prosodic boundary size. While the plosives following a Minor boundary had an average closure of approximately 100 ms, those following only Word boundaries were on average even shorter than 70 ms. Likewise, the preceding vowel was about 20 % longer before Minor and even 30% longer before Major boundaries than before Word boundaries.

In contrast, VOT, the most important cue to the fortis/lenis distinction according to our analysis, showed only limited prosodic effects. The effect was restricted to fortis plosives and the effect size was 8%. Also on the burst intensity maximum, there was only little prosodically conditioned variation. While there was no effect for bilabials at all, the mean differences between neighboring prosodic categories ranged from 0.8 to 1.0 dB for alveolars and velars. Finally, also the following vowel duration showed just a minimal effect of prosody (differences of 2 to 5 ms between prosodic categories), and the effect was restricted to vowels following velars, possibly because these vowels were rather long in our experiment.

On glottal vibration, in contrast, there was a huge effect of prosodic position, and it is indeed a question how relevant glottal vibration is for the fortis/lenis distinction in German. Glottal vibration is considered a sufficient but not a necessary cue. This is supported by our data which shows that 42% of all lenis plosives were produced without any glottal vibration (with the Major condition already being excluded).

Our findings thus appear in line with the phonology-over-prosody hypothesis. Prosody affects especially those acoustic characteristics that do not contribute to phonological contrasts. As a consequence, prosody hardly affects the perceptual distinction between fortis and lenis plosives.

The question now is why the minimal prosodic effects that we observe for the relevant fortis/lenis cues appear to diminish the contrast in higher prosodic positions (which is the opposite of the predictions made by the Uniform Strengthening Account and the Feature Enhancement Account). So far, we have only considered the acoustic characteristics of the fortis and lenis plosives in the different prosodic conditions without taking into account whether listeners are able to hear the differences that we observed. Some of the acoustic correlates may not be well perceivable, especially after Word boundaries. For instance, glottal vibration and burst intensity may be masked by the preceding vowel in the short plosives following word boundaries. As a consequence, the acoustic variation in these cues may be unconstrained. This would explain the unexpected direction of some of the attested prosodic effects.

In conclusion, our data suggest that domain-initial strengthening affects especially those cues that are not highly relevant for signaling phonological contrasts (confirming the phonology-over-prosody hypothesis). The small prosodic effects that we do observe on highly relevant cues do not reflect domain-initial strengthening in the form of either the Uniform Strengthening or the Feature Enhancement Account.

References

Allen, J.S. and Miller, J.L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America* **106**, 2031-2039.

van Alphen, P.M. and Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics* **32**, 455-491.

Baayen, H. (2008). *Analyzing linguistic data. A practical introduction to statistics*. Cambridge: Cambridge University Press.

Bates, D.M. (2005). Fitting linear mixed models in R. *R News* **5**, 27-30.

Beckman, M.E., Edwards, J. and Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In Docherty, G.J. and Ladd, D.R. (Eds.) *Papers in Laboratory Phonology* **II** (pp.68-86). Cambridge: Cambridge University Press.

Beckman, M. and 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. *Glot International* **5**, 341-345.

Bombien, L., Mooshammer, C., Hoole, P., Kühnert, B. and Schneeberg, J. (2006). An EPG study of initial /kl/ clusters in varying prosodic conditions in German. *Proceedings of the 7th International Seminar on Speech Production,* Ubatuba, Brazil.

Butcher, A. (1977). Coarticulation in intervocalic voiceless plosives and fricatives in connected speech. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)* **8**, 154-212.

Byrd, D. and Choi, S. (2006). At the juncture of prosody, phonology, and phonetics? The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Proceedings of the 10th Conference on Laboratory Phonology,* Paris, France.

Chatterjee, S., Hadi, A.S. and Price, B. (2000). *Regression analysis by example*. New York: John Wiley and Sons.

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* **33**, 121-157.

Cho, T. and Jun, S-A. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistic Society* **36,** 31–44.

Cho, T. and Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics* **29**, 155-190.

Cho, T. and Keating, P. (2007). Effects of initial position versus prominence in English. *UCLA Working Papers in Phonetics* **106***,* 1-33.

Cho, T. and 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 of English stop consonants. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 2661-2664.

Cole, J., Kim, H., Choi, H. and Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* **35**, 180-209.

Crystal, T. H. and House, A. S. (1988). Segmental durations in connected speech signals: Current results. *Journal of the Acoustical Society of America* **83**, 1553–1573.

Dalgaard, P. (2002). *Introductory statistics with R*. New York: Springer.

Ernestus, M., Lahey, M. F. Verhees, F. and Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America* **120**, 1040-1051.

Farnetani, E. and Vayra, M. (1996). The role of prosody in the shaping of articulation in Italian CV syllables. *Proceedings of the 1st ESCA Workshop on Speech Production Modeling, 4thSpeech Production Seminar*, 9-12. Autrans, France.

Féry, C. (1993). *German Intonation Patterns*. Tübingen: Niemeyer.

Fischer-Jørgensen, E. (1976). Some data on North German stops and affricates. *Annual Report Institute for Phonetics University Copenhagen (ARIPUC)* **10**, 149-200.

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

Fougeron, C. (1998). *Variations articulatoires en début de constituants prosodiques de different niveaux en français*. PhD Dissertation, Université Paris III.

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

Fox, A. (1993). The prosodic hierarchy in German. In Flood, J.L., Salmer, P., Sayce, O. and Wells C. (Eds.), *Das unsichtbare Band der Sprache. Studies in German Language and Linguistic History in Memory of Leslie Seiffert* (pp. 455-481). Stuttgart: Heinz.

Fuchs, S. (2005). Articulatory correlates of the voicing contrast in alveolar obstruent production in German. *ZAS Papers in Linguistics (ZASPiL)* **41**.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German*. MPI Series in Psycholinguistics **7**. PhD Dissertation, Nijmegen.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. doi:10.1016/j.jml.2007.11.007

Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruens in German*. Amsterdam: Benjamins.

Jessen, M. (2001). Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. In Hall, T.A. (Ed.), *Distinctive feature theory* (pp. 237-294). Berlin: Mouton de Gruyter.

Jessen, M. and Ringen, C. (2002). Laryngeal features in German. *Phonology* **19**, 189-218.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology* **15**, 189–226.

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* **97**, 491–504.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* **60**, 286–319.

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

Kohler, K.J. (1977). The production of plosives. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)* **8**, 30-110.

Kohler, K.J. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica* **36**, 332-343.

Kohler, K.J. (1982). F0 in the production of lenis and fortis plosives. *Phonetica* **39**, 199-218.

Kohler, K.J. (1984). Phonetic explanation in phonology. The feature Fortis/Lenis. *Phonetica* **41**, 150-174.

Kohler, K.J. (1995[2)]. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

Kuzla, C., Cho, T., Ernestus and M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics* **35**, 301-320.

Kuzla, C., Ernestus, M. and Mitterer, H. in press. Compensation for assimilatory devoicing and prosodic structure in German fricative perception. *Papers in Laboratory Phonology* **X**.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* **29**, 3-11.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Peterson, G., Lehiste, I. (1960). Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America* **32**, 693–703.

Pierrehumbert, J. and Talkin, D. (1992). Lenition of /h/ and glottal stop. In Docherty, G.J. and Ladd, D.R. (Eds.). *Papers in Laboratory Phonology* **II** (pp. 90-117). Cambridge: Cambridge University Press.

de Pijper, J.R. and 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**, 2037-2047.

Piroth, H.G. and Janker, P.M. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* **32**, 81-109.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook* **3** (pp. 371-405). Cambridge: Cambridge University Press.

Shattuck-Hufnagel, S. and Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* **25**, 193-247.

Slis, I.H. and Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction II. *Language and Speech* **12**, 80-102, 137-155.

Tabain, M. (2003a). Effects of prosodic boundary on /aC/ sequences: acoustic results. *Journal of the Acoustical Society of America* **113**, 516–530.

Tabain, M. (2003b). Effects of prosodic boundary on /aC/ sequences: articulatory results. *Journal of the Acoustical Society of America* **113**, 2834–2849.

Ülzmann, N. (2007). *Eine experimentalphonetische Analyse des intervokalischen /t-d/ Kontrastes im Norddeutschen*. Unpublished M.A. Thesis, Kiel University.

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

Prosodic strengthening of German fricatives in duration and assimilatory devoicing

Chapter 3

Kuzla, C., Cho, T., and Ernestus, M. (2007). *Journal of Phonetics* **35**, 301-320.

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, showing 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, indicating the influence of prosodic structure on sandhi processes. Reduction in glottal vibration made lenis fricatives more fortis-like (/f, z/). 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 shorter (more lenis-like). As a consequence, 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.

Introduction

The fine phonetic details of segment realizations are determined both by prosodic structure (e.g., Byrd, Kaun, Narayanan and Saltzman 2000; Cho 2002; Fougeron 2001; Keating, Cho, Fougeron and Hsu 2003; Pierrehumbert and Talkin 1992; Shattuck-Hufnagel and Turk 1996; Wightman, Shattuck-Hufnagel, Ostendorf and Price 1992) and by sandhi processes such as coarticulation and assimilation (cf. Kühnert and 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 (cf. Nespor and 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 and Cooper 1975; Pirello, Blumstein and 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.

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 and Keating 1997; Cho and Keating 2001; Fougeron 2001), stops show longer closures and longer Voice Onset Time (Pierrehumbert and Talkin 1992; Jun 1998; Keating et al. 2003), and vowels are more often glottalized (Dilley, Shattuck-Hufnagel and 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 (Nespor and 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 Nespor and Vogel 1986:179):

(1) a) vocalic context: *les visiteurs sont* [sõt] *arrivés* 'the visitors have arrived',

b) citation form: *sont* [sõ] '(they) are',

c) consonantal context: *les visiteurs sont* [sõ] *partis* 'the visitors have left'.

d) Le garçon les aidait [ɛdɛ] // activement. 'The boy helped them actively'.

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 1990:79; 1995:160). 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.

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 and van Hout 2003 for Dutch; Cole and Cooper 1975; Pirello et al. 1997; Stevens et al. 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 and Vagges 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 and Cohen 1969a, 1969b; van den Berg and 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/cm^2) must be created between subglottal and supraglottal air pressure (Westbury and 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.

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 1998:177), 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 (1998:188) 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."

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 and Baumann 2000 for German). In contrast, the existence of a tonally marked intermediate phrase level is controversial in German (Féry 1993; Grice and Baumann 2000, Grabe 1998; Uhmann 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).

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 increases 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/.

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.

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.

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 I) 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 I. 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 I, Sentence). Such a boundary would be equivalent to the “utterance” boundary in Nespor 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 super ordinate clause’s verb and a following NP in an obligatory complement clause (Table I, 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 Nespor 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 I, 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 I). This resulted in non-accentuation of the target words.

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 treesdrawn 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 hollows and hills.'*
Word /v/:	**Benno** hat Wälder und Seen gemalt. Benno has forests and lakes drawn *'**Benno** has drawn hollows and hills.'*
Word /z/:	**Clara** hat Senken und Hügel gemalt. Clara has hollows and hills drawn *'**Clara** has drawn hollows and hills.*

Table 1. Sentence sets for the sequences /t#f/, /t#v/, and /t#z/. 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.').

The full crossing of experimental factors yielded 24 conditions, containing one test sentence each: 3 fricatives (f, v, z) x 2 preceding contexts (/t/ vs. /ə/) x 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 sentence types. The order of sentence types within blocks was pseudo-randomized and different for each block in a single list.

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.

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 ms 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 - 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 and 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 intonation phrase 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 II shows the token counts for prosodic categories by syntactic stimulus type for the analyzed data points (cf. section 2.6. below).

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*

Table II: Number of tokens by syntactic type and prosodic category

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[2]. 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.

[2] The 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 be also be influenced by voicing.

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).

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.

Results

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). Figure 1 shows the average duration of the preboundary syllables /hat/ and /tə/ as a function of prosodic category and fricative.

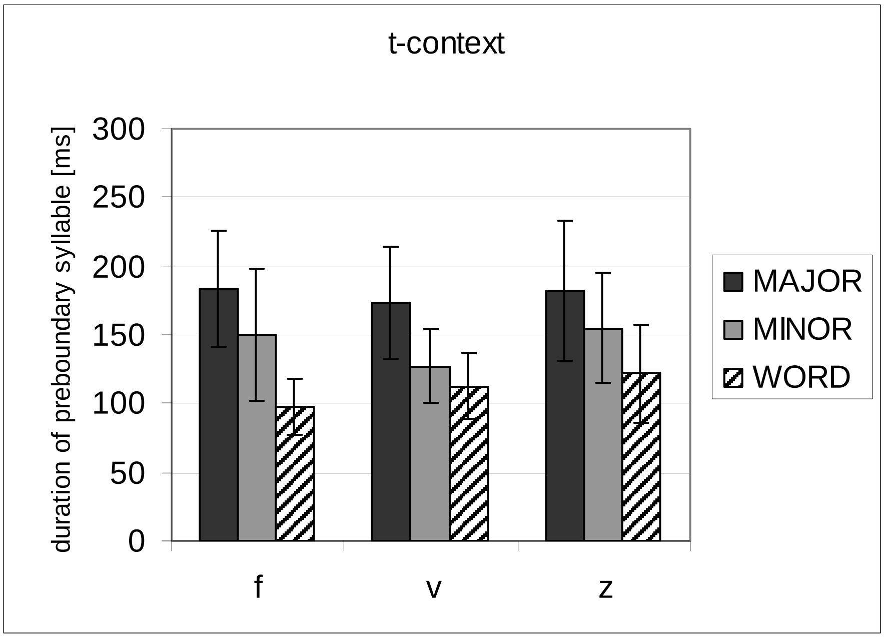


Figure 1a: Duration of the preboundary syllable in /t/-context as a function of postboundary fricative (f,v,z,) and prosodic boundary (MAJOR, MINOR, WORD).

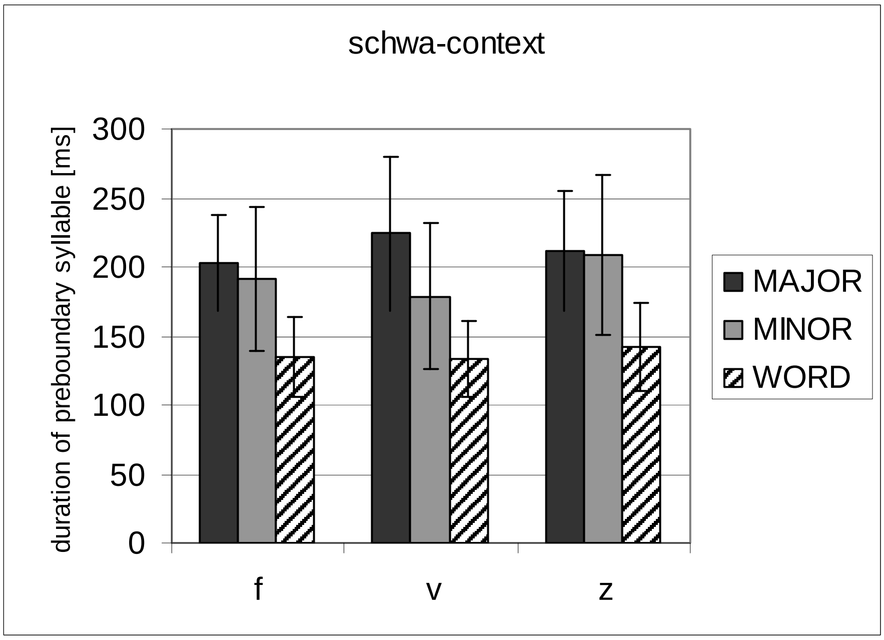


Figure 1b: Duration of the preboundary syllable in /ə/-context as a function of postboundary fricative (f,v,z,) and prosodic boundary (MAJOR, MINOR, WORD).

Here and for all other analyses reported in this study, we built multi-level models with Speaker as random variable (Pinheiro and Bates 2000; Baayen 2004; Quené and 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 and 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 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 intonation phrases in German.

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 ms 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. Figure 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 ms 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).

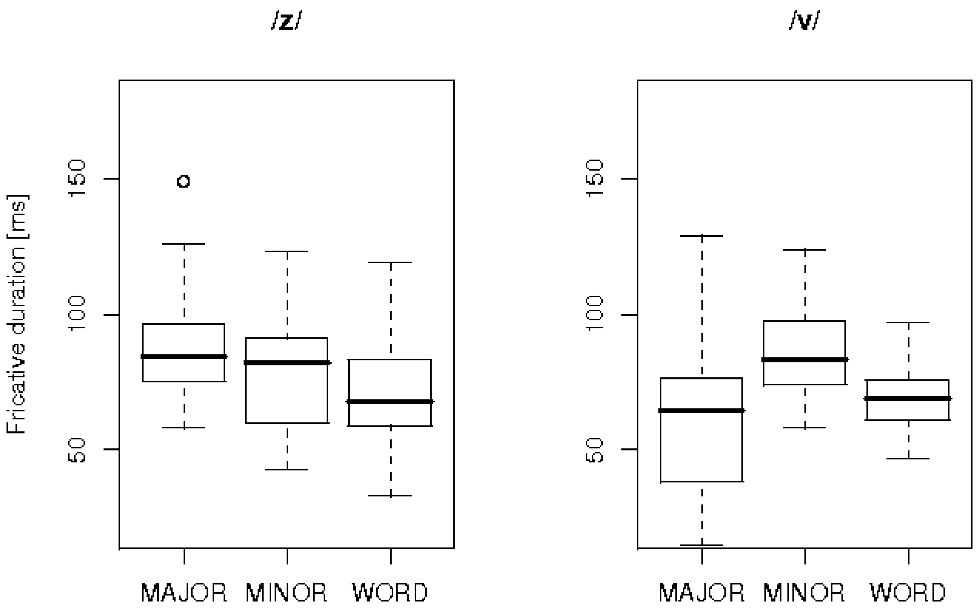


Figure 2: Distribution of the durations of /z/ and /v/ in /ə/-context as a function of prosodic boundary (MAJOR, MINOR, WORD).

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.

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 Figure 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 few milliseconds, which corresponds to a range from less than one up to approximately three or four glottal cycles, depending on the speaker's f0 and the given fricative duration. This accounts for why there are no tokens falling in the range of 91-99% in Figure 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.

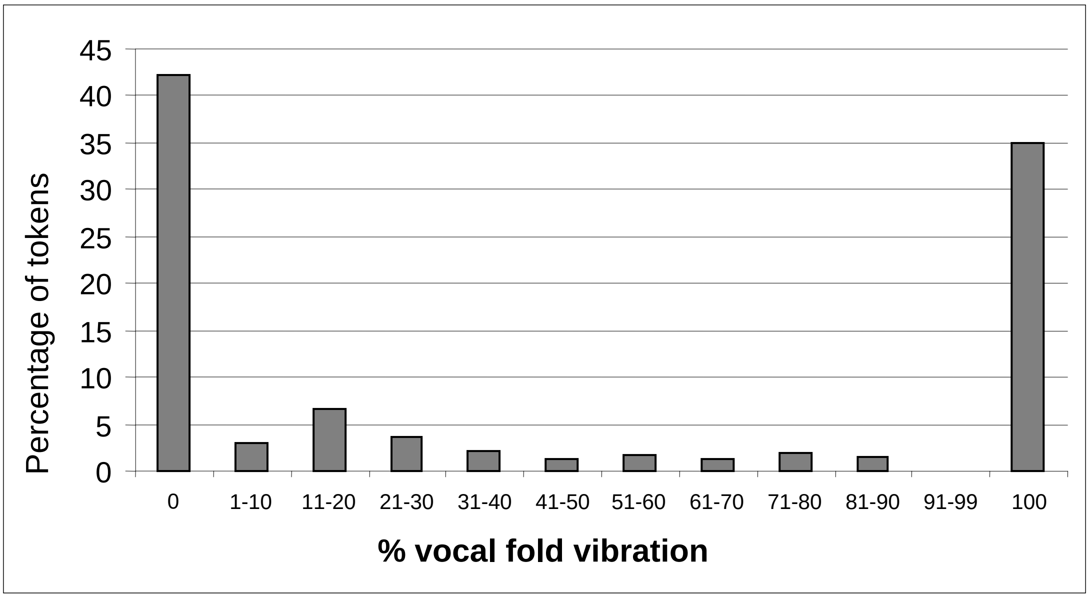


Figure 3: Distribution of percentages of vocal fold vibration.

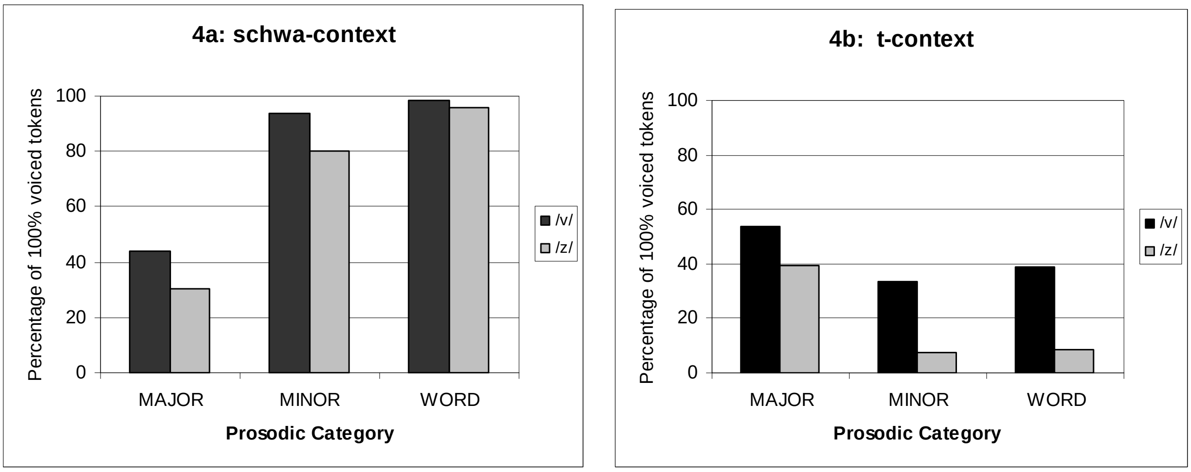
Likelihood of full voicing

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 Figure 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 Figure 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.



Figures 4a,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/).

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

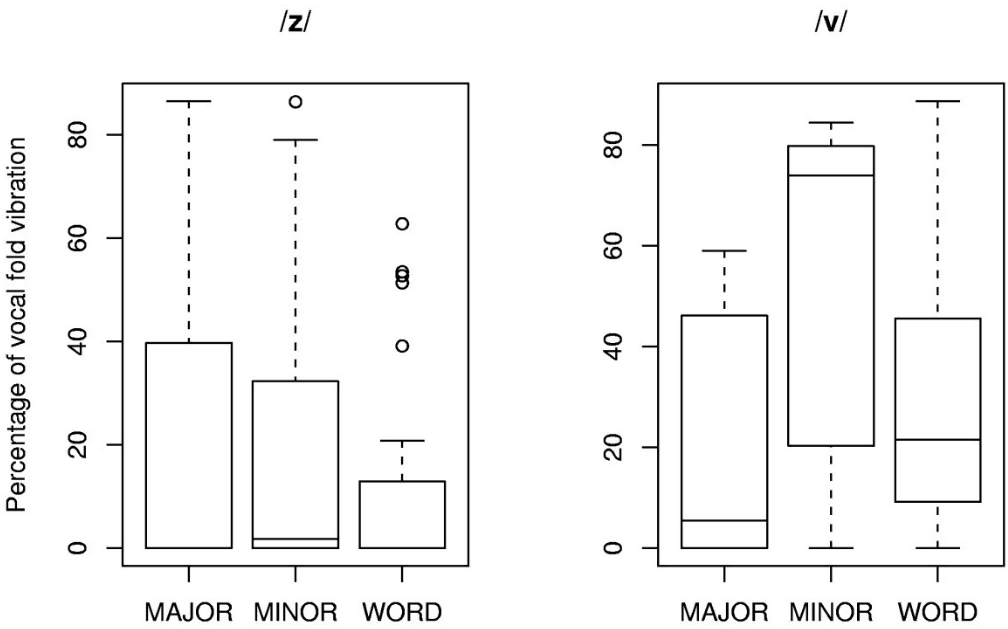


Figure 5: Distribution of the percentages of vocal fold vibration for /z/ and /v/ in /t/-context as a function of prosodic boundary (MAJOR, MINOR, WORD).

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 ($p < 0.05$).

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 (Figure 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 Figure 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.

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.

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/: $F(2,141) = 25.96$, $p < 0.001$). Figure 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 (97ms), 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 (Figure 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.

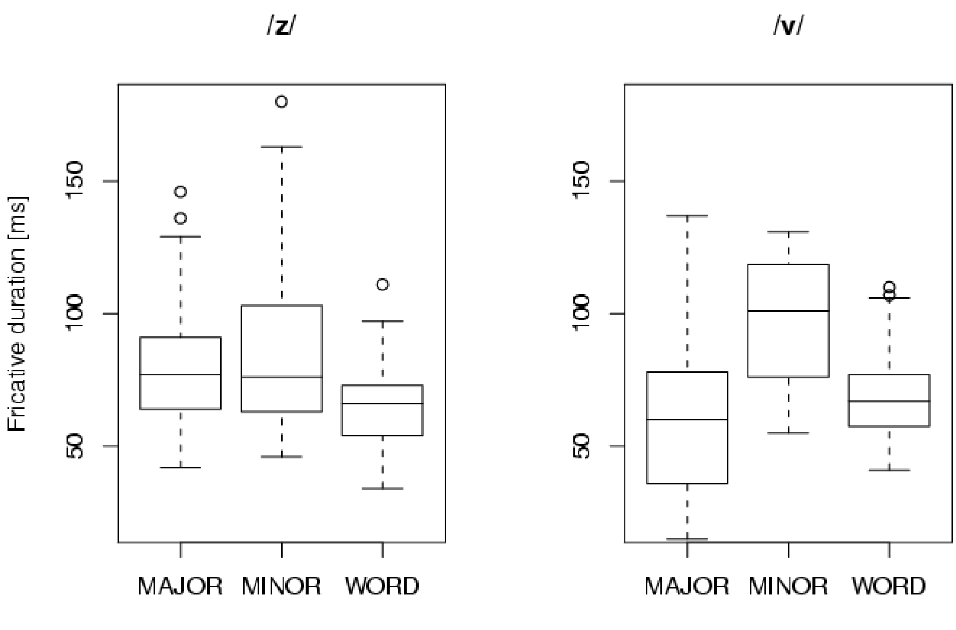


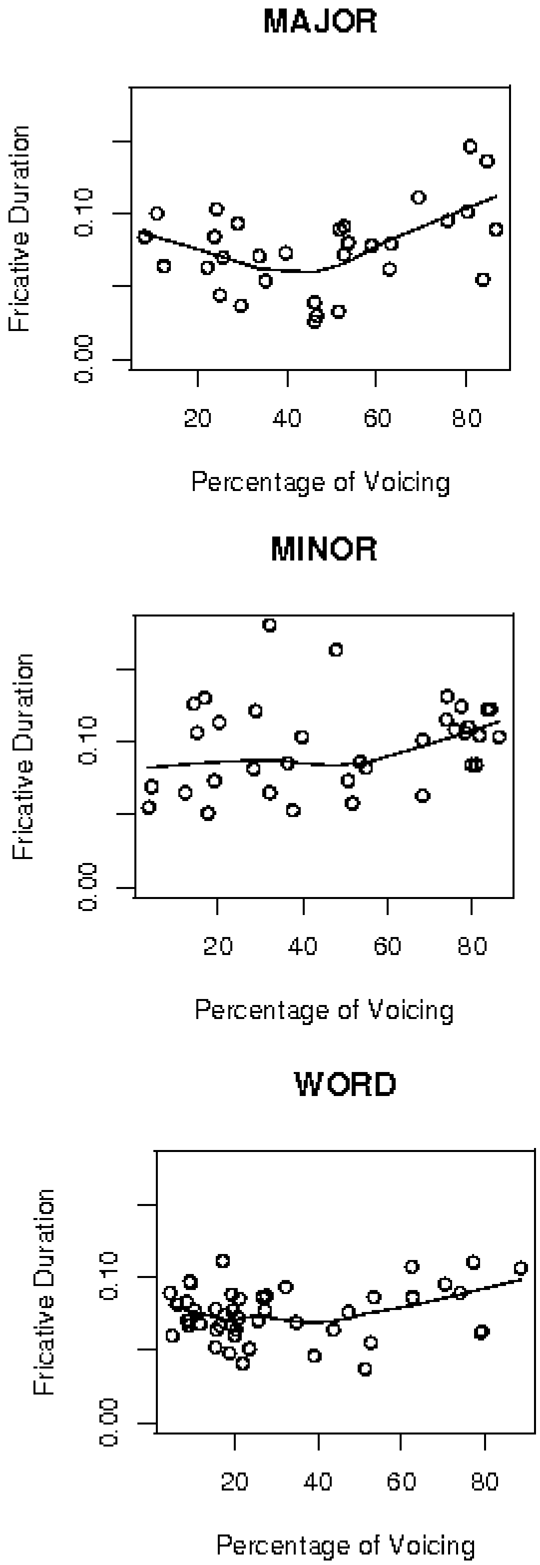
Figure 6: Distribution of the duration of /z/ and /v/ tokens in /ə/-context as a function of prosodic boundary (MAJOR, MINOR, WORD).

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. Figure 7a-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 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.



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

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.

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.

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 and Keating 2001, Keating et al. 2003 for Korean; Cho and McQueen 2005 for Dutch; Fougeron 2001, Keating et al. 2003, Tabain 2003a, b for French; Fougeron and 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; Figure 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 and 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 and 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. Whereas initiation of phonation may be delayed after the pause in the MAJOR condition, we can interpret this finding as an indirect result of initial strengthening in the MINOR condition. Prosodic strengthening induces lengthening of the fricatives, and it is more difficult to maintain frication and vocal fold vibration simultaneously over a longer period of time (Stevens et al. 1992).

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 and 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 and Cho 2005; Choi 2003; Cole et al. 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 and 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 and 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., Figure 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 and Nolan 1995; Nolan, Holst and 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 and 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.

References

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers* **1**, 1-45.

Baer, T. (1975). *Investigation of phonation using excised larynxes.* PhD dissertation, MIT.

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* **3**. Cambridge: CUP. 255-309.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International* **5**, 341-345.

Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon.

Browman, C. P and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica* **49**, 155-180.

Byrd, D., Kaun, A., Narayanan, S. and Saltzman, E. (2000). Phrasal signatures in articulation. In M. Broe and J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology* **V**. Cambridge: CUP. 70-87.

Byrd, D. and 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. and 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. and Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* **29**, 155-190.

Cho, T. and 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. *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona (Spain). 2662-2664.

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. and Hasegawa-Johnson, M. (2003). The effect of accent on the acoustic cues to stop voicing in radio news speech. *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona (Spain). 2665-2998.

Cole, R. A. and 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. and 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**. 2037-2047.

Dilley, L., Shattuck-Hufnagel, S. and 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: Fori.

Ernestus, M., Lahey, M., Verhees, F. and 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. and Vagges, K. (1979). Perceptual category shift of voiceless Italian fricatives as a function of duration shortening. In B. Lindblom and S. Öhman (Eds.), *Frontiers of speech communication research. Festschrift for Gunnar Fant*. London: Academic Press. 159-165.

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 Phonetics* **29**, 109-135.

Fougeron, C. and 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 and C. Wells (Eds.) *Das unsichtbare Band der Sprache. Studies in German language and linguistic history in memory of Leslie Seiffert.* Stuttgart: Heinz. 455-481.

Grabe, E. (1998). *Comparative intonational phonology: English and German.* MPI Series in Psycholinguistics 7. PhD Dissertation, Nijmegen.

Grice, M., and 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. and Nolan, F. (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell and A. Arvaniti (Eds). *Papers in Laboratory Phonology* **IV**. Cambridge: Cambridge University Press. 315-333.

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. and 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. and Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden and R. Temple (Eds.) *Papers in Laboratory Phonology* **VI.** Cambridge: Cambridge University Press. 145-163.

Kissine, M., van de Velde, H., and 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 and A. Marchal (Eds., *Speech production and speech modelling*. Dordrecht: Kluwer. 69-92.

Kohler, K. J. (1995^2). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

Kühnert, B. and F. Nolan (1999). The origin of coarticulation. In W.J. Hardcastle and N. Hewlett, *Coarticulation: Theory, data and technique*s. Cambridge: Cambridge University Press. 7-30.

Mitterer, H. (2003). *Understanding "gardem bench". Studies on the perception of assimilated word forms.* PhD Dissertation, Maastricht.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nolan, F. Holst, T. and Kühnert, B. 1996. Modelling [s] to [ʃ] accomodation in English. *Journal of Phonetics* **24**, 113-137.

Peperkamp, S. (1997). *Prosodic words.* HIL Dissertations **34**. PhD Dissertation, Amsterdam.

Pierrehumbert, J. and Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty and D. R. Ladd (Eds., *Papers in Laboratory Phonology* **II**. Cambridge: CUP. 90-117.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* Statistics and Computing. New York: Springer.

Pirello, K., Blumstein, S. E. and Kurowski, K. (1997): The characteristics of voicing in syllable-initial fricatives in American English. *Journal of the Acoustical Society of America* **101**, 3754-3765.

Quené, H. and van den Bergh, H. (2004): On multi-level modeling of data from repeated measures designs: A tutorial. S*peech Communication* **43**, 103-121.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook* **3**. Cambridge: CUP. 371-405.

Shattuck-Hufnagel, S. and 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. and Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing,* 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. and Cohen, A. (1969a). On the complex regulating the voiced-voiceless distinction I. *Language and Speech* **12**, 80-102.

Slis, I. H. and 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. and 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 /aC/ sequences: acoustic results. *Journal of the Acoustical Society of America* **113**, 516-530.

Tabain, M. (2003b). Effects of prosodic boundary on /aC/ sequences: articulatory results. *Journal of the Acoustical Society of America* **113**, 2834-2849.

Uhmann, S. (1991). *Fokusphonologie: eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie.* Tübingen: Niemeyer.

van den Berg, R. J. H. and 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. and 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. and 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.

Compensation for assimilatory devoicing and prosodic structure in German fricative perception

---

Chapter 4

Kuzla, C., Ernestus, M., and Mitterer, H. (in press). *Papers in Laboratory Phonology* **X**

Abstract

Two important sources of phonetic variation in German fricatives are progressive voice assimilation and prosodic structure: the lenis fricatives /v/, /z/ are devoiced after /t/ across word boundaries. This process is gradient and moderated by the size of the prosodic boundary before the assimilated word-initial fricative. Fricatives are more devoiced after smaller prosodic boundaries. In this study, we present three phoneme identification experiments. Fully voiced, partially devoiced and completely devoiced fricatives had to be identified as fortis or lenis in different segmental (assimilation vs. non-assimilation context) and prosodic (after a word vs. a phrase boundary) conditions. 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 the lexicon: we found prosodic effects on compensation for the lexically functional /f-v/ contrast, but not for the non-functional /s-z/ contrast. These findings suggest that a prosodic analysis of spoken language contributes to the resolution of lexical ambiguity arising from assimilation.

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, Hallé, and Segui 2006; Kuzla, Cho, and Ernestus 2007). 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. (2007) 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 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, Csépe, 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. 2007). 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. (2007) 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. (2007) 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. (2007) 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 Bölte 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. (2007) 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.

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 co-vary 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.

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.

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.

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). 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).

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 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.

For the word boundary continuum, we selected a recording of the sentence *Anna hatte Wälder und Wiesen gemalt* where the initial /v/ in *Wälder* was produced with continuous glottal vibration (f0 = 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 Wälder 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.

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).

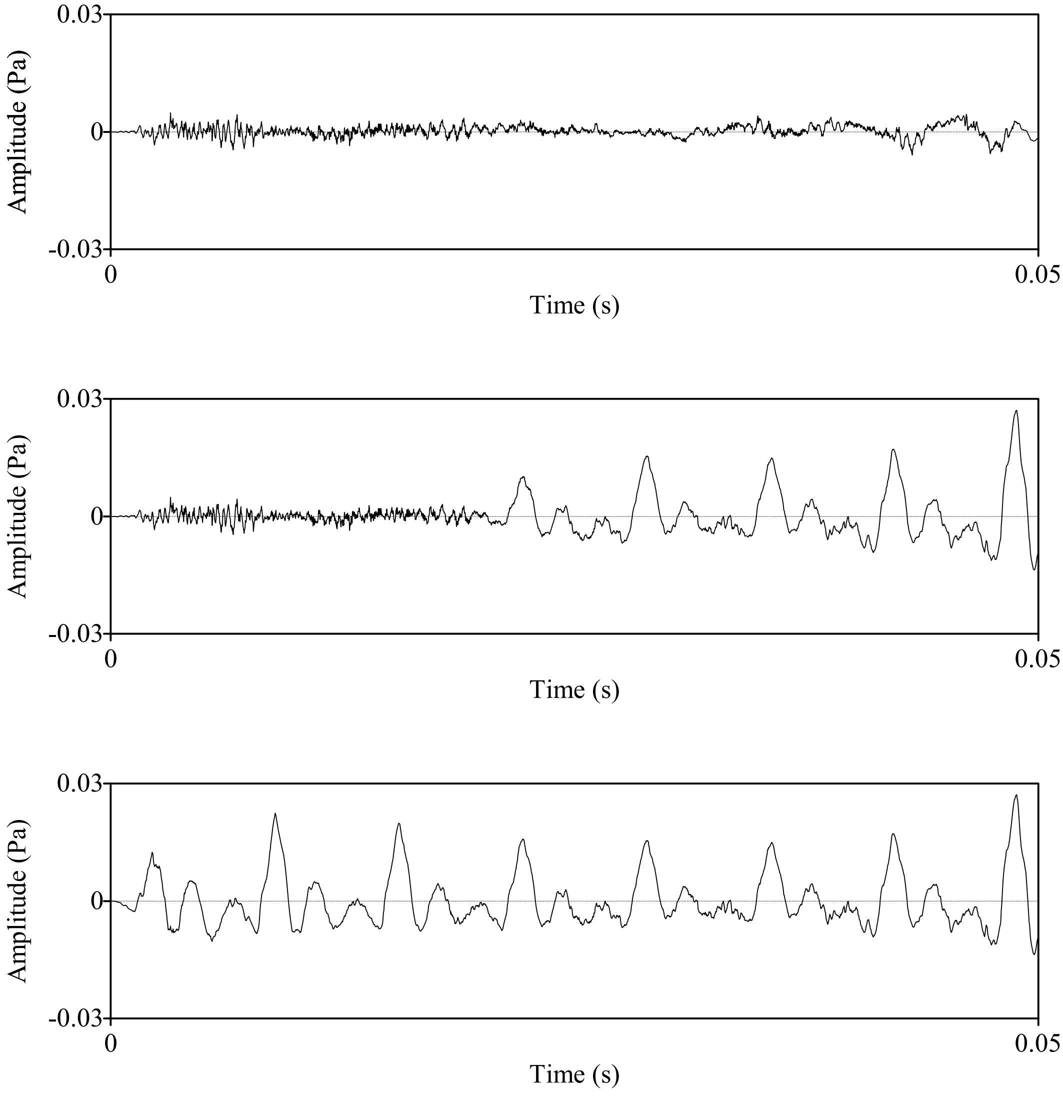


Figure 1: Acoustic waveform of the steps 1, 4 and 7 from the [v̥ - v] continuum for the prosodic word condition in Experiment 1

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 disproportionally 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% (±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 and Liberman, 1987).

The test sounds were spliced into the four sentences which contained the target word *Wälder* 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.

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 *Wälder* on the screen.

Participants were instructed that they would hear sentences containing manipulated versions of the words *Felder* and *Wälder*, 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 *Wälder* by pressing the response button labeled ‘F’ for *Felder* and the response button labeled ‘W’ for *Wälder*. 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.

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*2*7*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 ($F(6, 90) = 0.682$, $p = 0.67$).

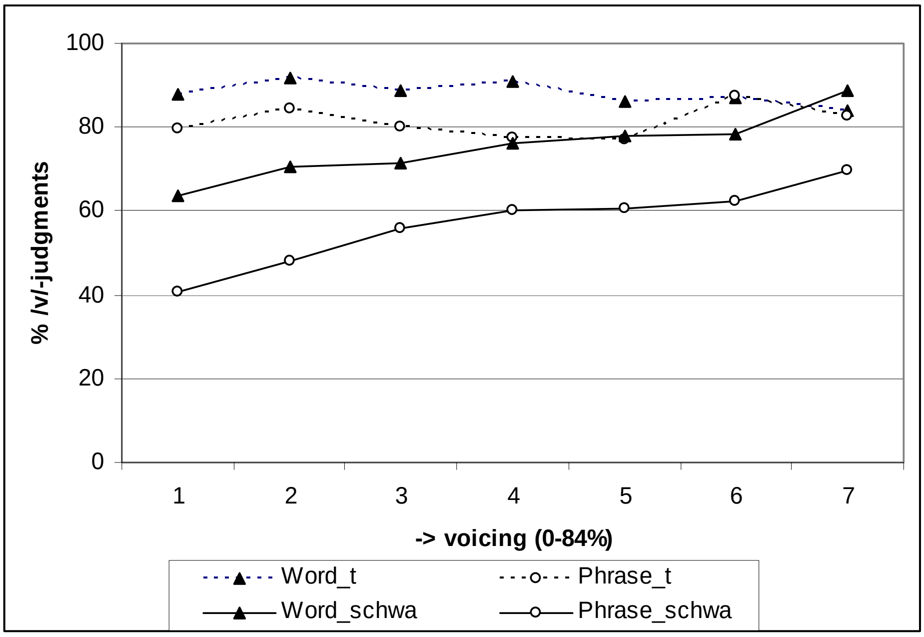


Figure 2: Percentages of /v/-judgments across the [v̥ - v] continua in the four prosodic+segmental context conditions in Experiment 1

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% to76%, 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. 2007). 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.

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.

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.

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.

Phrase Boundary, Assimilation Context
Weil sie vorhat, <u>Felder</u> 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 fieldss and meadows, she goes to the countryside.'
Phrase Boundary, Non-assimilation Context
Weil sie vorhatte, <u>Felder</u> 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 <u>Felder</u> und Wiesen gemalt.
Anna has fields and meadows painted
'*Anna has painted fields and meadows.'*
Word Boundary, Non-assimilation Context
Anna hatte <u>Felder</u> und Wiesen gemalt.
Anna had fields and meadows painted
'*Anna had painted fields 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.

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).

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.

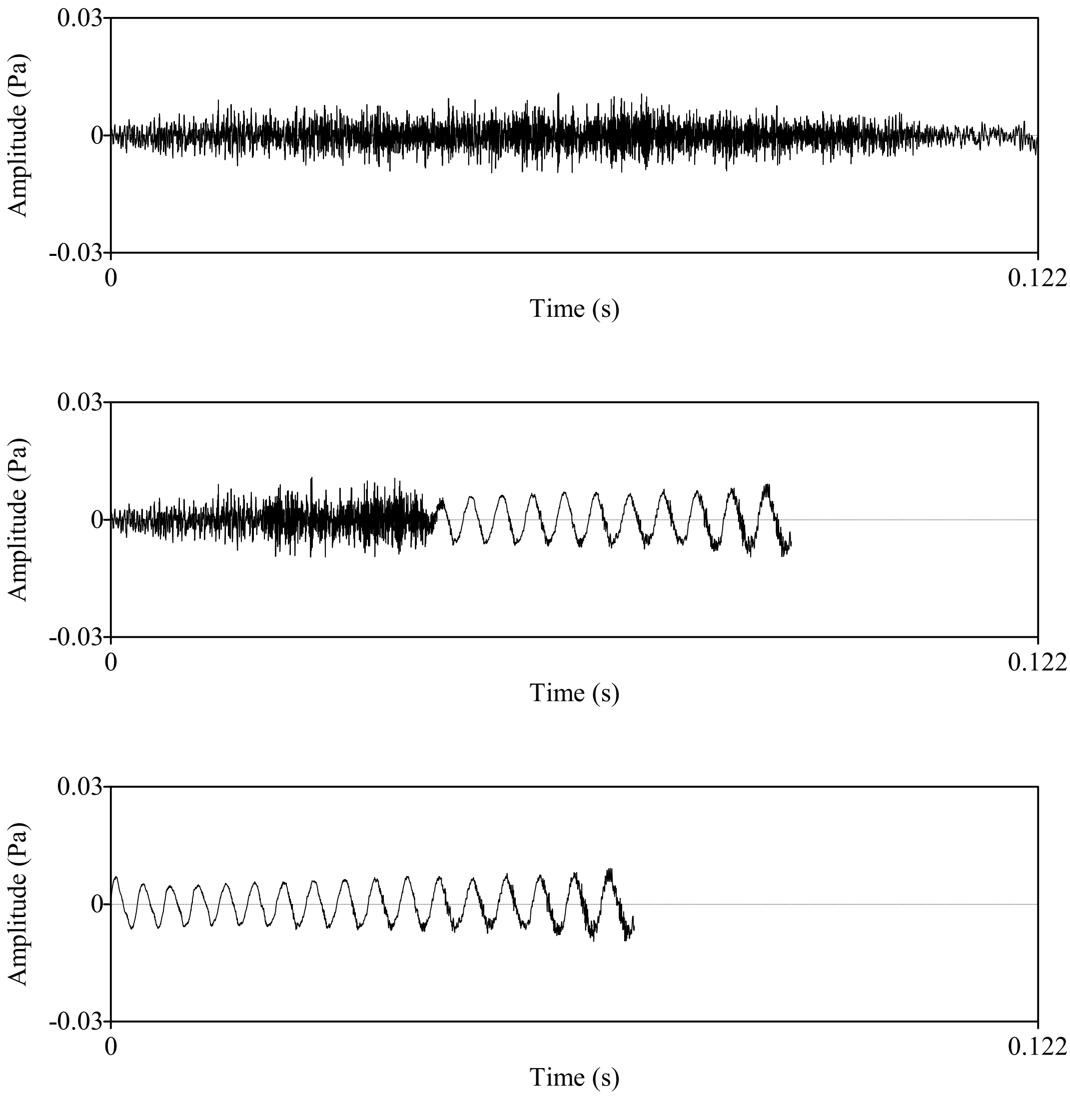


Figure 3: Acoustic waveforms of the steps 1, 10 and 20 from the /f - v/ continuum in Experiment 2

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.

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.

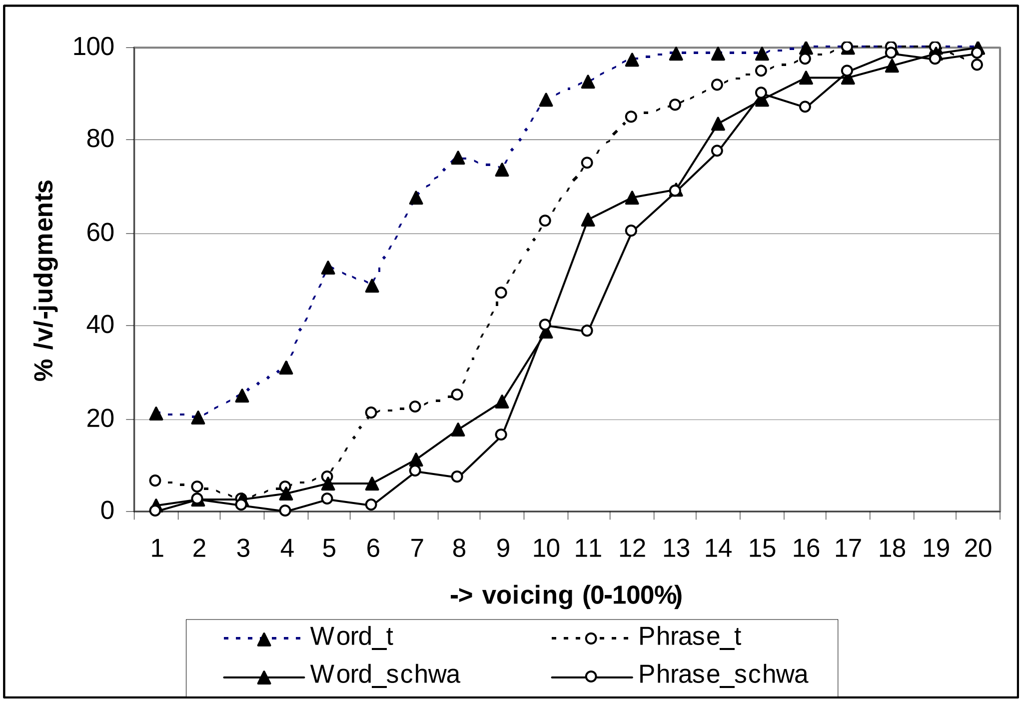


Figure 4: Percentages of /v/-judgments across the /f - v/-continuum in the four prosodic+segmental context conditions in Experiment 2.

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/).
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 *Wälder,* 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.

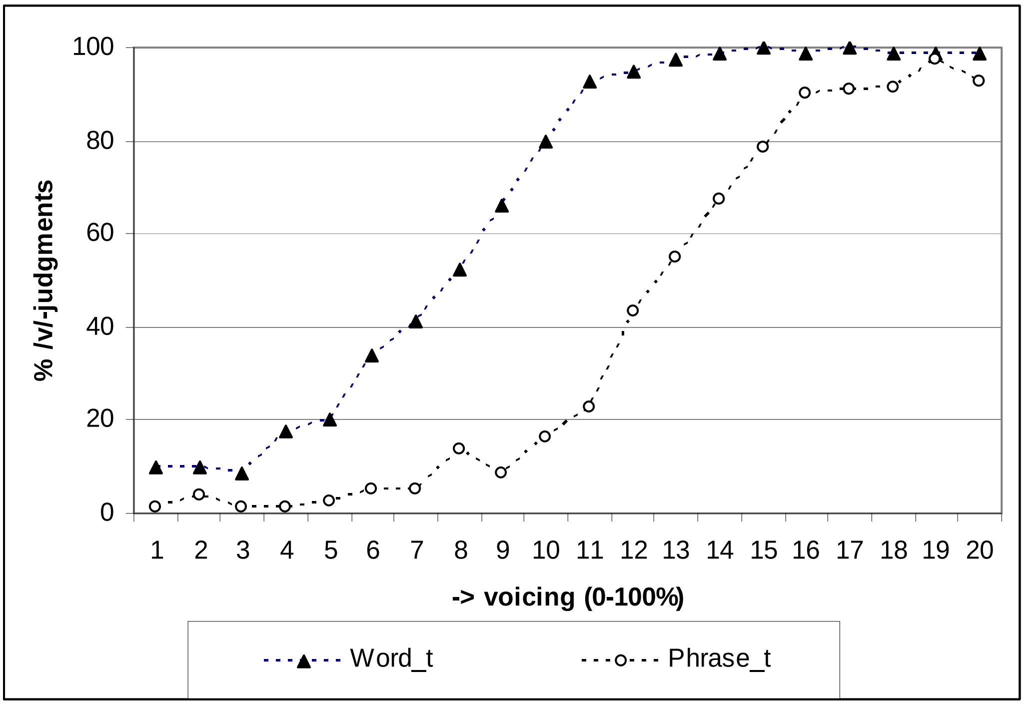


Figure 5: Percentages of /v/-judgments across the /f - v/-continuum in two prosodic conditions in the assimilation context, carrier sentences containing "Felder" in Experiment 2.

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.

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/.

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.

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 /ə/ plus Word boundary, and 4. Non-assimilation context /ə/ plus Phrase boundary. The materials were recorded in the same session as the materials for Experiment 2 and by the same female speaker.

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.'*

Table 3: Speech materials for Experiment 3. The /z/-initial target word was "Senken"(underscored). A contrastive accent on another word, 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".

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.

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/.

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 ($F(19, 361) = 6.994$, $p < 0.001$).

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.

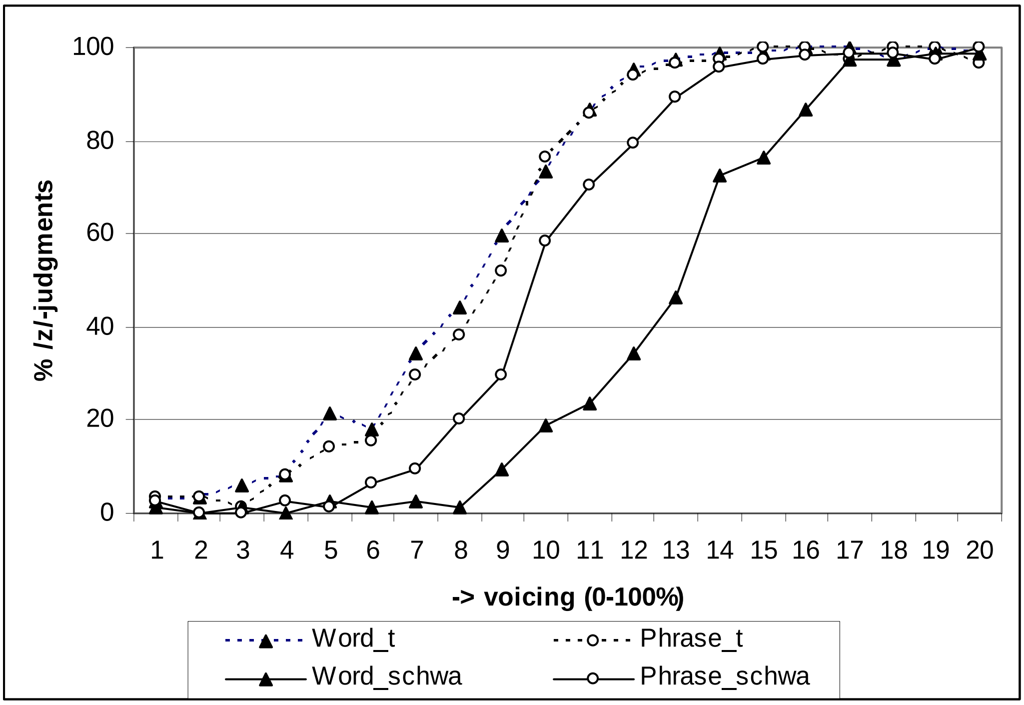


Figure 6: Percentages of /z/-judgments across the /s - z/-continuum in the four prosodic+segmental context conditions in Experiment 3.

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. 2007). 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.

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. (2007) 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 *Wälder – Felder* [vɛldɐ] - [fɛldɐ] 'forests' – 'fields', *Vase-Phase* [vazə] - [fazə] 'vase' – 'phase', and *wischen- fischen* [vɪʃən] - [fɪʃən] '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 (i.e., word-initial /v/ versus /f/) to non-functional (i.e., word-initial /z/ versus /s/) 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.2007) 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/, and is also more common for /z/ than for /v/ (Kuzla et al. 2007). 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 [hatv̥ɛldɐ] 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, and 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 [hatv̥ɛldɐ] 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.

References

Andruski, J. E., Blumstein, S.E., and Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition* **52**, 163 - 187.

Carlson, K., Clifton, C. Jr., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* **45**, 58 - 81.

Cho, T. and Keating, P. A. (2001). Articulatory and acoustic studies on domain initial strengthening in Korean. *Journal of Phonetics* **29**, 155 - 190.

Cho, T. and 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.

Cho, T., McQueen, J. M., and Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* **35**, 210 - 243.

Coenen, E., Zwitserlood, P. and Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* **16**, 535 - 564.

Darcy, I. (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, I., Peperkamp, S., and Dupoux, E. (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, M., Lahey, M., Verhees, F. and Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America* **120**, 1040 - 1051.

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

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

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

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

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

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

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

Keating, P. A., Cho, T., Fougeron, C. and Hsu, C.-S. (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, R. J., Ernestus, M., Schreuder, R. and Baayen, R. H. (2005). Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory and Cognition* **33**, 430 - 446.

Kjelgaard, M. M. and Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* **40**, 153 - 194.

Kohler, K. 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, K. J. (1995^2). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

Kuzla, C., Cho, T. and Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics* **35**, 301-320.

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

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology* **15**, 189-226.

Mitterer, H. (2003). Understanding “gardem bench”. Studies on the perception of assimilated word forms. PhD Dissertation, Maastricht.

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

Mitterer, H., Csépe, V. and Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology* **59**, 1305 - 1334.

Mitterer, H. ,Csépe, V., Honbolygo, F., and Blomert, L. (2006). The recognition of assimilated word forms does not depend on specific language experience. *Cognitive Science* **30**, 451–479.

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

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition* **52**, 189 - 234.

Repp, B. H. and Liberman, A. M. (1987). Phonetic categories are flexible. In Harnad, S. (Ed.), *Categorical Perception: The groundwork of cognition.* Cambridge, Mass.: Cambridge University Press. 89 - 112.

Salverda, A. P., Dahan, D. and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* **90**, 51 - 89.

Snoeren, N. D., Hallé, P. A. and Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* **34**, 241 - 268.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M. and Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. Journal of the Acoustical Society of America **91**, 2979 - 3000.

Perceptual compensation for voice assimilation of German fricatives

Chapter 5

Kuzla, C., Mitterer, H., and Cutler, A. (2006). *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, Auckland, New Zealand, December 6 – 8

Abstract

In German, word-initial lax fricatives may be produced with substantially reduced glottal vibration after voiceless obstruents. This assimilation occurs more frequently and to a larger extent across prosodic word boundaries than across phrase boundaries. Assimilatory devoicing makes the fricatives more similar to their tense counterparts and could thus hinder word recognition. The present study investigates how listeners cope with assimilatory devoicing. Results of a cross-modal priming experiment indicate that listeners compensate for assimilation in appropriate contexts. Prosodic structure moderates compensation for assimilation: Compensation occurs especially after phrase boundaries, where devoiced fricatives are sufficiently long to be confused with their tense counterparts.

Introduction

Assimilation contributes substantially to the variance in spoken language. As a consequence, it causes a potential challenge for word recognition, since listeners must map different word forms to the same lexical representation. There are several solutions to this problem. First, it is known that listeners compensate for assimilation by adapting their phoneme categories to segmental contexts. In contexts that license assimilation (viable contexts), phonological feature changes in the assimilated speech sounds are easily tolerated, whereas such changes seriously hinder the recognition of the intended word in environments that do not license the assimilation (non-viable contexts; e.g., Gaskell and Marslen-Wilson 1996; Coenen, Zwitserlood and Bölte 2001; Gow 2003; Mitterer and Blomert 2003). Second, assimilation is often gradient rather than categorical, whereby the acoustic details of assimilated word forms preserve cues to the underlying form. Listeners take advantage of such fine phonetic detail (Gow 2003; Nolan 1992; Snoeren, Hallé and Segui 2006). Finally, assimilation is constrained by prosodic structure: Larger prosodic boundaries inhibit the occurrence and degree of assimilation (Nespor and Vogel 1986; Kuzla, Cho and Ernestus 2007), which is in line with phonetic research on the articulatory strengthening of segments in prominent prosodic positions, in particular at the beginning of prosodic domains (Keating, Cho, Fougeron and Hsu 2003). A prosodic analysis of the spoken utterance may help listeners to identify a word form as assimilated or not.

In German, fricatives can be devoiced after voiceless obstruents, even across word boundaries (Kohler 1995; Jessen 1998). As a result, assimilated fricatives can become more similar to their voiceless counterparts, leading to increased ambiguity for minimal pairs such as [vaɪn] *Wein* ‘wine’ and [faɪn] *fein* ‘fine’. Kuzla, Cho and Ernestus (2007) showed that this assimilation is graded and conditioned by prosodic structure, such that fricatives are devoiced more often and to a greater extent after prosodic word boundaries than after prosodic phrase boundaries. Importantly, even complete devoicing does not lead to neutralization of the tense-lax contrast, since other cues to this contrast are not affected. For instance, a completely devoiced /v/ is still differentiated from /f/ by its shorter duration.

Despite the overall maintenance of phonological contrasts, the question remains whether listeners are sensitive to the systematic variation in glottal vibration, one of the major cues to laxness. Compensation for assimilation predicts that a devoiced /v/ should be perceived as less similar to /f/ in assimilation contexts than in non-assimilation contexts. Moderation of this compensation by prosodic structure would imply that a devoiced /v/ is perceived as more similar to /f/ after phrase boundaries than after word boundaries, because assimilatory devoicing is less frequent and often incomplete in this prosodic position. Using an off-line phoneme categorization task, Kuzla, Ernestus and Mitterer (in press) found that listeners' compensation for assimilatory devoicing is indeed moderated by these prosodic effects if other cues (e.g., duration, frication energy) are covaried with the amount of glottal vibration during the fricative. If these secondary cues remain consistent with the lax fricative /v/, there is generally little confusability with /f/. Nevertheless, compensation for assimilation still occurs: more /v/-responses are obtained in assimilation context than in non-assimilation context. This context effect is larger after a prosodic phrase boundary, where fricatives are longer than they are after a word boundary. This suggests that the durational cue to laxness preserved in the fine phonetic detail of the assimilated fricative is more informative in the Word condition than in the Phrase condition.

The goal of the present study is to further investigate the role of phonetic detail in combination with segmental and prosodic context in the perception of devoiced fricatives. In contrast to Kuzla, Ernestus and Mitterer (in press), we use an online task, in order to get better insight into the early processes of compensation for assimilation.

Experiment

We used cross-modal identity priming to examine this issue. In this task, participants have to decide whether a visually presented target word is an existing word or not (lexical decision) shortly after they have heard an auditory stimulus (the prime). Reaction times are typically faster if the target word has been heard shortly before. Responses are slower if a competing word has been presented. If one member of a minimal pair is used as the target, this task allows us to investigate whether an auditorily presented form is perceived as different or similar to that member, as a function of segmental and prosodic contexts.

The compensation for assimilation hypothesis predicted that a devoiced /v/ in /vaɪn/ *Wein* 'wine' primes /faɪn/ *fein* 'fine' in non-assimilation context, where devoiced /v/ typically does not occur and therefore may be interpreted as /f/. In contrast, devoiced /v/ should not prime, or even inhibit, /faɪn/ in assimilation context, since devoiced /v/ may be the assimilated form of /v/ from /vaɪn/ (Compensation for Assimilation Hypothesis). Second, we expected these differences in priming to be stronger after a prosodic phrase boundary than after a word boundary, because complete devoicing is less natural after a Phrase boundary in production, and because other phonetic cues to /v/ (e.g., duration) may be less salient in this prosodic position, so that listeners have to rely on context more (Prosodic Structure Hypothesis).

Materials and design

Fifty minimal pairs of German mono- and bisyllabic words were selected (see Appendix). The members of the pairs differed in the phonological voicing of their initial labiodental fricative, as in /vaɪn/ *Wein* ‘wine’ versus /faɪn/ *fein* ‘fine’. The prime words appeared in semantically neutral sentence contexts, which biased neither the /v/- nor the /f/- initial words. The /f/-initial members of the minimal pairs served as targets in the visual lexical decision task, and also as prime words in the Identity condition. The corresponding /v/-initial words were the auditory prime words in our four experimental conditions. For these conditions, the context before the prime words (referred to as the pre-prime in the following) contained the manipulation of two experimental factors: Immediately preceding segmental context (/ə/ versus /t/; i.e., viable versus non-viable assimilation context), and prosody (Word boundary versus Phrase boundary). The crossing of these two factors Context and Prosody yielded four experimental conditions (Word_e, Word_t, Phrase_e, and Phrase_t). In the Word_e condition, the prime word was preceded by /ə/ from *hatte* ‘had’ (non-viable assimilation context), and a prosodic word boundary (Example 1). The Word_t condition was the same except that the preceding segment was /t/ from *hat* ‘has’ (viable assimilation context, triggering devoicing). In the Phrase_e condition, the preceding segment was /ə/ from *vorhatte* ‘wanted’, but the prosodic boundary was a phrase boundary. Phonetically, the phrase boundary was characterized by preboundary lengthening and an intonation rise (‘high boundary tone’). In the fourth condition, Phrase_t, the context was /t/ from *vorhat* ‘(she) wants’ before a Phrase boundary (Example 2).

(1) *Anna hatte Wein richtig geschrieben.*
‘Anna had written wine correctly’.

(2) *Weil sie vorhat, Wein richtig zu schreiben, fragt sie Paul.*
‘Since she wants to write wine correctly, she asks Paul.’

Fifty phonologically unrelated words were selected as prime words for a control condition (the Unrelated condition). In addition, the experiment contained 150 fillers of three types, 50 of each type.

For filler Type A, prime words were 25 /v/-initial and 25 /f/-initial words which do not form minimal pairs with existing /f/- respectively /v/-initial words (e.g., *Waffel* ‘wafer’, **Faffel; Fabel* ‘fairy tale’, *Wabel). The targets were 50 non-words. Primes in the fillers of Type B were phonologically unrelated words, and targets were again non-words. Type C fillers consisted of phonologically unrelated prime words and unrelated existing target words. Due to this structure of the test items and fillers, the total numbers of Yes/No responses in the lexical decision task were equal. For the Identity and the Unrelated condition and for all fillers, the characteristics of the carrier sentences varied between the four segmental and prosodic conditions used in the experimental conditions. Table 1 provides an overview of the stimulus structure for all conditions and fillers with examples.

In addition, 20 catch trials were randomly inserted in the experiment to ensure that participants paid attention to the auditory stimuli. In these trials, participants had to answer a Yes/No question about the prime sentence (e.g., ‘Was it Anna who wrote wine correctly?’) instead of performing lexical decision on the target. For this purpose, the proper names in the carrier sentences were varied between eight bisyllabic female names in the pre-primes of the Word conditions and eight monosyllabic male names in the post-primes of the Phrase conditions.

<table>
<tr><th>Condition</th><th>Pre-prime</th><th>Prime word</th><th>Post-prime</th><th>Target</th></tr>
<tr><td>Identity</td><td rowspan="2">varying</td><td>fein
‘fine’</td><td rowspan="2">varying</td><td rowspan="6">fein
‘fine’</td></tr>
<tr><td>Unrelated</td><td>Biene
‘bee’</td></tr>
<tr><td>Word_e</td><td>Anna hatte_</td><td rowspan="4">Wein
‘wine’</td><td rowspan="2">_richtig geschrieben</td></tr>
<tr><td>Word_t</td><td>Anna hat_</td></tr>
<tr><td>Phrase_e</td><td>Weil sie vorhatte,_</td><td rowspan="2">_richtig zu schreiben, fragt sie Paul</td></tr>
<tr><td>Phrase_t</td><td>Weil sie vorhat,_</td></tr>
<tr><td>Filler
Type A
Type B
Type C</td><td>varying</td><td>Waffel
‘wafer’
(*Faffel)
Biene
‘bee’</td><td>varying</td><td>Non-word
Hand ‘hand’</td></tr>
</table>

Table 1: Stimulus structure

A female native speaker of German read several repetitions of the materials in a sound-attenuated booth. Recordings were made with a Sennheiser microphone, and directly stored onto a computer, at a sample rate of 48 kHz. From these recordings, we selected one realization of

- each pre-prime carrier sentence,
- each /f/-prime word,
- each unrelated prime word,
- each filler prime word,
- each post-prime carrier sentence,
- each /v/-prime word from a Word_e context, and
- each /v/-prime word from a Phrase_e context.

What is important to note here is that we selected two prosodically different realizations of the /v/-prime words because the acoustic details of domain-initial words are known to be affected by the prosodic hierarchy (e.g., Keating et al. 2003).

To investigate the perceptual effects of devoicing, the initial fricatives of all /v/-prime words within the same prosodic category were replaced by an identical acoustic sound. For the Word conditions, we resynthesized a completely voiceless /v/ produced by our speaker in assimilation context during the recording. This sound had a duration of 50 ms. The sound used in the Phrase conditions was created by expanding the duration of the sound from the Word conditions to 70 ms. These two durations are typical for the two prosodic domains for our speaker. The speech editing was done with the PSOLA component of the PRAAT software package (Boersma 2001). The selected recordings of the pre-primes, the prime words and the post-primes were combined online during the experiment to form the auditory stimuli of the respective conditions.

In order to ensure sufficient statistical power despite the limited number of items, two versions of the experiment were created. Context was treated as a within-subjects factor, and Prosody as a between-subjects factor. Version 1 of the experiment contained the Word_e condition, the Word_t condition, the Identity condition, and the Unrelated condition. Version 2 contained the Phrase_e condition, the Phrase_t condition, the Identity, and the Unrelated condition. Moreover, each version of the experiment contained an additional condition testing for factors not reported in the present paper. The priming conditions in both versions were counterbalanced across five different lists, and each participant saw each target only once. Each list contained 224 trials: four training trials, the 50 experimental trials, 150 fillers, and 20 catch trials.

Participants

One hundred native listeners of German participated in the experiment for a small payment. Most of them were students at Kiel University, and none of them reported any hearing disorders. Half of them received the Word version of the experiment, and the other half the Phrase version.

Procedure

Participants were tested one by one in a quiet room by means of a portable computer. They were instructed to listen to the auditory primes and to decide as quickly as possible whether the stimulus appearing on the computer screen was an existing word or not. Furthermore, participants were told that occasionally they would have to answer a written question about the sentence they had just heard. Participants indicated their lexical decisions and answers to the questions by pressing one out of two response buttons. The experiment was controlled by NESU (=Nijmegen Experimental Set-Up) software. The auditory primes were presented at a comfortable listening level via headphones. The visual targets appeared in a large font on the computer screen half a prime word duration after offset of the prime word, i.e., during the post-prime part of the auditory prime. This ensured that listeners had heard the prosodic information spread over the pre-prime and the prime word when the target appeared. After each block of 50 trials participants were allowed to take a short break. An experimental session lasted about 30 minutes.

Results and discussion

Mean reaction times of the correct responses to the /f/- initial target words per condition are shown in Table 2. RTs below 400 ms and above 1500 ms were discarded. For all analyses, we fitted linear mixed-effects (lme) models with Participant and Item as random factors (e.g., Pinheiro and Bates 2000; Baayen 2004). RTs were logarithmically transformed into lnRT to obtain a quasi-normal distribution of this dependent variable.

First, we examined the conditions that appeared in both versions of the experiment, that is, the Identity condition and the Unrelated condition. An analysis of lnRTs in these two conditions with Condition and Experiment Version as fixed factors yielded a significant effect of condition ($F(1, 1713) = 33.30$, $p < 0.001$), that is, facilitation in the Identity condition. There was neither a main effect of Experiment Version nor an interaction of Condition with Experiment Version. These results allowed us to pool the data from the two experimental groups for further analysis.

Condition (prime)	Mean RT [ms]
Identity	794
Unrelated	832
Word_e (non-assim.)	847
Word_t (assim.)	837
Phrase_e (non-assim.)	833
Phrase_t (assim.)	859

Table 2: Mean reaction times in ms per condition

To examine the effects of segmental context and prosodic structure on the perception of devoiced fricatives, we restricted the dataset to the experimental conditions (Word_e, Word_t, Phrase_e and Phrase_t). We analyzed lnRT as function of Prosody (Word, Phrase), Context (e, t), and Trial Number. Trial Number was included in the model as a covariate, since participants might have become faster or slower during the experiment as a consequence of practice or fatigue. Trial Number was significant (faster responses with increasing Trial Number, $F (1, 1607) = 156.59, p < 0.001$), as it was in both of the following separate analyses, but there were no significant interactions involving this factor. Neither Prosody ($F (1, 1607) = 0.10, p = 0.75$) nor Context ($F (1, 1607) = 1.19, p = 0.28$) emerged as significant; however, there was a significant interaction of these two factors ($F (1, 1607) = 6.11, p < 0.05$). In order to investigate the interaction, we split the data by Prosody and modeled the lnRT as a function of Context and Trial Number. After a Word boundary, there was no effect of context ($F (1, 849) = 0.48, p = 0.49$). In contrast, after a Phrase boundary, Context was significant ($F (1,758) = 4.47, p < 0.05$), indicating faster RTs in the non-assimilation (/ə/-) context than in the assimilation (/t/-) context.

Finally, posthoc Bonferroni comparisons were carried out between the four experimental conditions and the unrelated conditions in the respective versions of the experiment. These comparisons revealed a significant difference between the Phrase_t condition and the unrelated condition, indicating inhibition, whereas the other conditions did not differ from the Unrelated condition. These findings are in line with the difference found between Phrase_t and Phrase_e in the previous analysis. In addition, posthoc comparisons with Bonferroni adjustments showed that the Identity condition differed significantly from all experimental conditions (all $p < 0.001$). The priming effects can be seen in Figure 1. The pattern shows that a devoiced /v/ never leads to equally fast responses as the identity prime-word starting with /f/, but that, surprisingly, only in Phrase_t, there is significant inhibition.

These results partly confirm our hypotheses. As predicted by the Compensation for Assimilation Hypothesis, there is an effect of context, but only after a Phrase boundary. However, this is evidence for the expected difference between the prosodic conditions (Prosodic Structure Hypothesis).

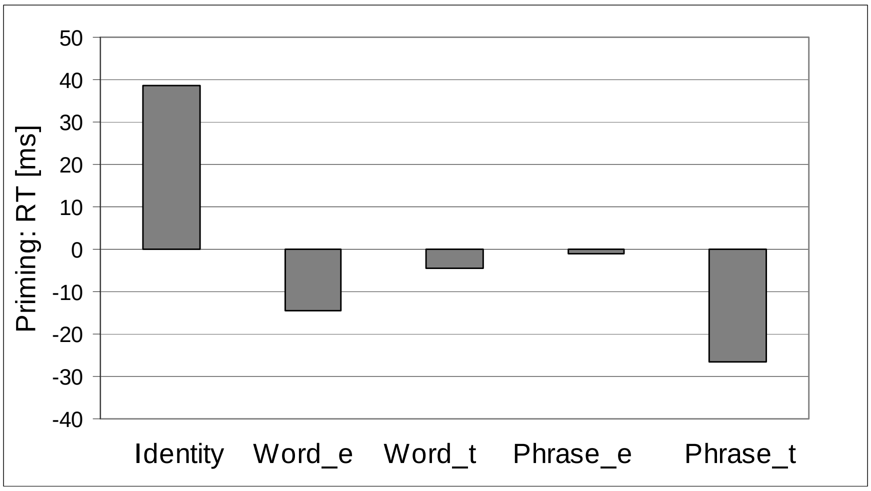


Figure 1: Priming results computed as the difference between RTs in the conditions on the x-axis and in the unrelated control condition, with a positive value indicating facilitation and a negative inhibition

General Discussion

This study investigated the role of context and prosodic boundary size in the perception of devoiced fricatives at word onsets in German. Lax fricatives such as /v/ can be devoiced after voiceless obstruents (Kohler 1995, Jessen 1998). This assimilation occurs more frequently and to a larger extent across smaller prosodic boundaries (Kuzla, Cho and Ernestus, 2007). Devoicing affects an important cue to the phonological tense-lax contrast in minimal pairs such as /vaɪn/ *Wein* ‘wine’ and /faɪn/ *fein* ‘fine’.

We examined how segmental context and prosodic structure influenced the perception of a devoiced fricative as /v/ or /f/ in a cross-modal priming experiment with the /f/-initial words as targets. Responses to these targets should be faster the more similar an auditory prime is to the target.

A study by van Alphen and McQueen (2006) applied a similar method investigating the role of prevoicing in Dutch word-initial stops. These authors found that making a lax stop (normally prevoiced) more tense-like by removing the prevoicing influences the lexical activation of competing words starting with a voiceless plosive. Similarly, Andruski, Blumstein and Burton (1994) reported in a study on the effects of fine phonetic detail (Voice Onset Time) of English stops that phonologically similar competitor word candidates are activated in a graded fashion, depending on the subphonemic acoustic shape of the prime words.

In our experiment, as expected, there was significant identity priming, i.e., facilitation if the /f/-initial word had been heard as a prime. We hypothesized that the segmental context would influence the perception of a devoiced fricative, i.e., that it would be recognized as /v/ more easily in assimilation that in non-assimilation context. This hypothesis was partly confirmed, as the predicted effect was observed after a Phrase boundary, but not after a Word boundary. The difference between the two prosodic boundary conditions, however, confirms our second hypothesis that compensation for assimilation would be modulated by prosodic structure. After a prosodic word boundary, we did not find any effect of context, neither priming nor inhibition. In contrast, after a phrase boundary, there was inhibition in assimilation context, but not in non-assimilation context. We take this as evidence for compensation for assimilation. However, there was again no priming in the non-assimilation context. The lack of priming in assimilation context after both prosodic boundaries indicates that there are always sufficient secondary cues that distinguish a devoiced /v/ from /f/. After word boundaries, these other cues (e.g., duration) leave no role for compensation for assimilation. After phrase boundaries, the fricative is longer than after word boundaries, and this longer duration appears to increase the ambiguity between a devoiced /v/ and /f/; hence listeners take the context into account and compensate for assimilation. In the non-assimilation context, the mismatch between the segmental environment and the voicing characteristics of the fricative in the prime leads to faster responses to the /f/-targets than in the assimilation context. Apparently, there is less competition from the prime. That the reaction times are nevertheless not as fast as if the listeners had heard a genuine /f/ suggests that the prime is perceived as intermediate between /v/ and /f/. Together, the results of our study indicate that segmental context, fine phonetic detail and prosodic structure interact in a complex way in on-line recognition of spoken words.

References

Andruski, J. E., Blumstein, S. E. and Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition* **52**, 163 -187.

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers* **1**, 1- 45.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International* **5**, 341-345.

Coenen, E., Zwitserlood, P. and Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* **16,** 535-564.

Gaskell, M. G. and Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology; Human Perception and Performance* **22**, 144-158.

Gow, D.W. Jr. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics* **65**, 575-590.

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

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

Kohler, K. J.(1995^2). *Einführung in die Phonetik des Deutschen.* Berlin: Erich Schmidt.

Kuzla, C., Cho, T. and Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics* **35**, 301-320.

Kuzla, C., Ernestus, M. and Mitterer, H. (in press). Compensation for assimilatory devoicing and prosodic structure in German fricative perception. *Papers in Laboratory Phonology* **X**.

Mitterer, H. and Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics* **65**, 956-969.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

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

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S_PLUS.* (Statistics and Computing). New York: Springer.

van Alphen, P. M. and McQueen, J.M. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology; Human Perception and Performance* **32**,178-196.

Snoeren, N. D., Hallé, P. A. and Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* **34**, 241 - 268.

Appendix: Minimal Pairs

v-prime words		**f-prime words/targets**	
wach	*awake*	Fach	*partition*
Wachmann	*guard*	Fachmann	*expert*
wackeln	*to shake*	fackeln	*to hesitate*
Waden	*calves*	Faden	*thread*
wäre	*were*	Fähre	*ferry*
Werte	*values*	Fährte	*track*
wellig	*wavy*	fällig	*due*
Welt	*world*	fällt	*falls*
werben	*to advertise*	färben	*to colour*
wahren	*to keep*	fahren	*to drive*
warten	*to wait*	Fahrten	*rides*
Wall	*dam*	Fall	*fall*
Walter	*(proper name)*	Falter	*butterfly*
Wand	*wall*	fand	*found*
Wangen	*cheeks*	fangen	*to catch*
warm	*warm*	Farm	*farm*
was	*what*	Fass	*barrel*
wachsen	*to grow*	faxen	*to fax*
Wächter	*guard*	Fechter	*fencer*
weder	*neither*	Feder	*feather*
wehen	*to blow*	Feen	*fairies*
Wege	*ways*	fege	*sweep*
Wähler	*voter*	Fehler	*error*
Weiher	*pond*	Feier	*party*
Weile	*while*	Feile	*file*
Wein	*wine*	fein	*fine*
Wälder	*forests*	Felder	*fields*
Welle	*wave*	Felle	*furs*
Werner	*(proper name)*	ferner	*further*
Weste	*waistcoat*	Feste	*festivities*
wetten	*to bet*	fetten	*to grease*
wetzen	*to sharpen*	Fetzen	*rag*
Wichte	*wights*	Fichte	*spruce*

Wiese	*meadow*	fiese	*nasty (pl)*
winden	*to wind*	finden	*to find*
winken	*to wave*	Finken	*finches*
wischen	*to wipe*	fischen	*to fish*
Wort	*word*	fort	*away*
Wrack	*wreck*	Frack	*dress-coat*
Wuchs	*growth*	Fuchs	*fox*
wühlen	*to dig*	fühlen	*to feel*
Wunde	*wound*	Funde	*findings*
Vase	*vase*	Phase	*phase*
Weilchen	*while dim.*	Veilchen	*violet*
Wetter	*weather*	Vetter	*cousin*
wir	*we*	vier	*four*
Wolke	*cloud*	Volke	*people (infl)*
Wolle	*wool*	volle	*full (infl)*
Wok	*wok*	Fock	*foresail*
wedeln	*to fan*	fädeln	*to thread*

(dim. = diminutive; infl. = inflected form)

Summary and conclusions

Chapter 6

The relative articulatory strength of a speech sound depends partly on its position in the utterance. For example, sounds at the beginning of a word are pronounced more clearly than word-medial and word-final sounds, and they are diachronically more stable. Numerous studies in various languages (e.g., Cho and Keating 2001; Fougeron 2001; Keating et al. 2003) have also shown that this type of articulatory strengthening not only occurs at the beginning of words, but also in larger prosodic domains, such that phrase-initial sounds are articulated more strongly than word-initial, but phrase-medial, ones. In addition, irrespective of the prosodic position, different speech sounds may contrast with each other in 'strength'; this is true, for instance, of /p/ versus /b/ in German. Such a phonological contrast is traditionally known as a fortis-lenis contrast, and fortis-lenis distinctions between consonant classes are common in many languages. 'Articulatory strength' can thus result from two different sources: on the one hand from membership in a phonologically 'strong' sound category (i.e., fortis), and on the other hand from a prosodically strong position.

The objective of this thesis was to shed light on the interplay of the phonological fortis-lenis distinction and prosodic strengthening. In particular, I investigated how domain-initial prosodic strengthening affects the acoustic characteristics that are assumed to cue phonological fortis-lenis contrasts between German plosives (Chapter 2). Chapter 3 addressed the question how prosodic structure constrains a sandhi process, the progressive voice assimilation of fricatives, and explores the consequences of assimilatory devoicing for the fortis-lenis distinction. In addition, I examined the role of prosodic structure in speech perception. Chapters 4 and 5 reported experiments to investigate whether listeners take prosodically-conditioned variation into account when they have to categorize incoming speech sounds as fortis or lenis during spoken word recognition, or to activate word forms in memory.

Prosodic structure and phoneme contrasts in speech production

The experiments reported in Chapters 2 and 3 addressed the relationship of prosodic strengthening and the phonological fortis-lenis contrast in speech production. Chapter 2 focused on plain domain initial strengthening of German plosives. Speakers read various sentences which contained occurrences of the fortis plosives /p,t,k/ and the lenis plosives /b,d,g/ in word-initial position, in minimal word pairs. The sentences were constructed to elicit different-sized prosodic boundaries before the target words, and a *post hoc* prosodic analysis of the recordings confirmed that they indeed served this purpose. Acoustic measurements of various well-established cues to the fortis-lenis distinction revealed a complex relationship between prosodic strengthening and phonological contrast. In contrast to previously reported findings from other languages (e.g., Cho and Jun 2000 for Korean; Cho and McQueen 2005 for Dutch), not all cues were affected by prosody in the way suggested by the two hitherto proposed accounts of prosodic strengthening, the Uniform Strengthening account and the Feature Enhancement account. Uniform strengthening would imply that all plosives became more fortis-like in stronger prosodic positions, whereas Feature Enhancement predicts that fortis plosives display more fortis characteristics, too, but lenis plosives would become more lenis-like. However, there was no support for either account in the German data. Rather, the data support a contrast-over-prosody principle (see also Cole et al., 2007), stating that prosodic strengthening mostly affects those acoustic properties which do not serve as primary phonetic cues to phonemic contrasts.

A similar result was obtained in the study on the progressive assimilatory devoicing of fricatives reported in Chapter 3. In the German fricative inventory, there is a fortis-lenis contrast between /f/ versus /v/ and between /s/ versus /z/. However, in the native vocabulary, /s/ does not occur in word initial-position. The lenis fricatives /v/ and /z/ are assimilated to preceding voiceless obstruents, that is, they become devoiced. This process applies even across word boundaries. An acoustic analysis showed that assimilatory devoicing is gradient and conditioned by the size of the prosodic boundary between the obstruent and the fricative. Across prosodic word boundaries, there was less glottal vibration, that is, more assimilation than across phrase boundaries. This result is in line with previous research which has shown that prosodic strengthening implies greater resistance to coarticulation (Cho 2004).

Importantly, /z/ was generally more devoiced than /v/, a finding that can be explained by the fact that /v/ has a fortis counterpart /f/ in word-initial position, whereas devoicing of /z/ does not lead to potential phonemic confusion, because of the lack of word-initial /s/. Importantly, the duration of the fricatives was also affected by prosodic structure. Longer durations were found after phrase boundaries than after word boundaries. In other words, the greater amount of assimilatory devoicing which would make fricatives more confusable with their fortis counterparts at word boundaries was counteracted by the short duration, which was actually too short to let the devoiced fricatives appear fortis-like. In sum, there was again no uniform direction of the prosodic effects on different cues to the phonological contrast.

To conclude, both the plosive and the fricative study showed independent and often complementary effects of prosody on individual acoustic cues to the fortis-lenis contrasts, such that the contrasts were always preserved. Some of our findings also support the view recently proposed by Cole and colleagues (Cole et al. 2007) that prosodic strengthening mainly operates on those acoustic characteristics of speech sounds that do not primarily cue phonological contrasts.

Prosodic structure and compensation for assimilation in speech perception

The production studies summarized in the previous section suggest that, despite the systematic prosodic effects on some acoustic characteristics of speech sounds, prosodic strengthening never creates lexical ambiguities. Therefore, it is questionable whether listeners are exploiting this variation in word recognition. Chapter 4 reported three phoneme identification experiments which were designed to further explore the findings on assimilatory devoicing of fricatives in speech production, as described in Chapter 3.

Listeners were presented with test words whose initial sounds were replaced with sounds from resynthesized continua ranging from fully voiced, via partially devoiced through completely devoiced fricatives. The task was to identify these stimuli as fortis or lenis fricatives in different segmental environments (assimilation vs. non-assimilation context) and prosodic conditions (after a word boundary versus a phrase boundary). The results showed that listeners compensate for assimilatory devoicing, as they do for other types of assimilation (e.g., Coenen et al. 2001; Gow 2001, Mitterer and Blomert 2003). In the experiments presented in Chapter 4, participants judged partially devoiced fricatives more often to be lenis in the assimilation context than in the non-assimilation context. In addition, listeners appeared sensitive to prosodic structure when compensating for assimilation: more devoiced fricatives were more often judged as lenis after word boundaries than after phrase boundaries in assimilation context. Finally, there was also a lexical constraint on the effects of prosodic structure. A prosodic effect on the compensation for assimilation was attested for the lexically functional /f-v/ contrast, but not for the non-functional /s-z/ contrast. These findings suggest that a prosodic analysis of spoken language plays a role in the resolution of lexical ambiguity arising from assimilation.

There are two methodological limitations to the experiments reported in Chapter 4. First, phoneme categorization may capture only late stages of spoken language processing. Second, due to the large number of continuum steps and repetitions, the test sounds were tested in only one minimal word pair per fricative contrast., A further investigation of the perceptual compensation for assimilatory devoicing and its prosodic modulation was therefore undertaken using the cross-modal priming method (Chapter 5). In this task, participants have to perform a lexical decision task on a visually presented target word, that is, they have to indicate by pressing a button whether the target is an existing word or not, directly after having heard an auditory stimulus (the prime). If the target word has been heard shortly before, reactions to the targets are typically faster, whereas responses are slowed down if a lexical competitor word has been activated by the auditory stimulus. By using one member of a minimal pair as the target – in our experiment, the /f/-initial words -, it can be determined whether a stimulus just heard before is perceived as different or similar to that target, and whether the priming effects are affected by the segmental and prosodic context in which the primes occur. This experiment generally confirmed the results from the phoneme categorization experiments. After prosodic phrase boundaries, listeners compensated for assimilatory devoicing in the licensing segmental environment. Devoiced fricatives slowed down reaction times to the /f/-initial target words in assimilation context, indicating that they were perceived as good realizations of the /v/-initial competitor words. In a context that did not license assimilatory devoicing, these sounds were perceived as ambiguous between /v/ and /f/. The stimuli at the prosodic word boundaries may have been too short for the emergence of the effect. However, this difference between the prosodic conditions is in itself remarkable, as it confirms the hypothesis that prosodic structure modulates compensation for assimilation.

Future directions

While the results reported in Chapter 4 clearly indicated that listeners' compensation for assimilation was modulated by the prosody-driven degree of assimilation, this effect emerged only in case of the /f/-/v/ contrast. No such effect could be observed for the non-distinctive /s/-/z/opposition. The question remains open whether this difference is to be explained by a phonological, or a lexical constraint. For the both phonologically legal and lexically existing distinction in *Wälder* [vɛldɐ] 'forests' versus *Felder* [fɛldɐ] 'fields', listeners took the prosodic variation in the degree of devoicing into account. If the relevant level is the phonology, we would predict the same prosodic effect when the targets comprise the phonologically legal /f/-/v/ distinction even if tested on are a word and a nonword, for instance *Wasser* [vasɐ] 'water' versus *[fasɐ]. However, if the effects depends on the lexical status of the continuum endpoints, responses should pattern as in the *Senken* [zɛŋkən] 'hollows' versus *[sɛŋkən] experiment, that is, not display any effect of prosody. Such an investigation would give more insight into where in the speech perception system compensation processes take place, at a pre-lexical phonological level, or at the lexical level.

Broader implications

From the research presented in this thesis it has to be concluded that prosody and the segmental level are intertwined, but separate systems. From the production studies, it appeared that prosody influences the acoustic details of speech sounds in a systematic manner, but that variation is pretty much limited to acoustic dimensions that are not yet occupied by cueing phonological distinctions. However, by the employment of carefully controlled experimental settings it could be shown that listeners are amazingly sensitive to all the systematic variation in the speech signal and take all available information into account when performing a speech recognition task.

This thesis adds to previous work on the role of prosodically conditioned phonetic detail in speech production (e.g., Cho and Keating 2001, 2007; Fougeron and Keating 1997, Fougeron 2001; Keating et al. 2003, Tabain 2003a, b) and word segmentation in speech perception, (e.g., Salverda et al. 2003, Shatzman and McQueen 2006; Cho et al. 2007) and extends it to the phonemic identity of speech sounds.

References

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

Cho, T. and Jun, S-A. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistic Society* **36,** 31–44.

Cho, T. and Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* **29**, 155-190.

Cho, T. and McQueen, J.M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* **31**, 121-157.

Cho, T., McQueen, J. M., and Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* **35**, 210 - 243.

Coenen, E., Zwitserlood, P. and Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* **16**, 535 - 564.

Cole, J., Kim, H., Choi, H. and Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* **35**, 180-209.

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

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

Keating, P. A., Cho, T., Fougeron, C. and Hsu, C.-S. (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.

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

Salverda, A., Dahan, D. and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* **90**, 51-89

Shatzman, K. B. and McQueen, J. M. (2006) Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics* **68**, 1-16

Tabain, M. (2003a). Effects of prosodic boundary on /aC/ sequences: acoustic results. *Journal of the Acoustical Society of America* **113**, 516-530.

Tabain, M. (2003b). Effects of prosodic boundary on /aC/ sequences: articulatory results. *Journal of the Acoustical Society of America* **113**, 2834-2849.

Samenvatting en conclusies

De relatieve articulaire sterkte waarmee een spraakklank wordt geproduceerd is deels afhankelijk van de positie van de spraakklank in de zin. Klanken aan het begin van een woord worden duidelijker uitgesproken dan klanken in het midden of aan het eind van een woord, en ze zijn bovendien diachronisch gezien aan minder veranderingen onderhevig. Talloze onderzoeken in verscheidene talen (bijvoorbeeld Cho en Keating, 2001; Fougeron, 2001; Keating et al., 2003) hebben aangetoond dat dit soort articulaire versterking niet alleen aan het begin van woorden voorkomt, maar ook bij meer omvattende stukken spraak, de zogenaamde prosodische domeinen. Klanken aan het begin van een frase worden bijvoorbeeld sterker gearticuleerd dan klanken aan het begin van een woord die niet samenvallen met het begin van de frase. Onafhankelijk van hun positie in de prosodische structuur, kunnen verschillende spraakklanken bovendien onderling in 'sterkte' contrasteren; dit is bijvoorbeeld het geval voor de klanken /p/ ('sterker') en /b/ ('zwakker') in het Duits. Een dergelijk fonologisch contrast wordt ook wel een fortis-lenis-contrast genoemd, en fortis-lenis-verschillen tussen soorten medeklinkers komen in veel talen voor. Articulaire versterking kan dus veroorzaakt worden door twee verschillende processen: aan de ene kant doordat een klank behoort tot een groep klanken die sterker (oftewel fortis) worden gearticuleerd, aan de andere kant doordat een klank zich in een prosodische positie bevindt die versterkt wordt.

Het doel van dit proefschrift is om licht te werpen op de wisselwerking van het fonologische fortis-lenis-onderscheid en prosodische versterking. Meer in het bijzonder heb ik onderzocht hoe prosodische versterking aan het begin van een prosodisch domein de akoestische eigenschappen beïnvloedt waarvan vermoed wordt dat ze een aanwijzing zijn voor het fortis-lenis-onderscheid tussen Duitse plofklanken (zie Hoofdstuk 2). In Hoofdstuk 3 wordt de progressieve stemassimilatie van fricatieven bestudeerd. Een stemhebbende fricatief wordt stemloos wanneer deze een stemloze klank volgt. Ik heb onderzocht hoe de sterkte van de prosodische grens tussen deze twee klanken het assimilatieproces beïnvloedt, en wat de consequenties van stemassimilatie voor het fortis-lenis-onderscheid zijn. Ook heb ik de rol van prosodische structuur in spraakwaarneming onderzocht. In Hoofdstuk 4 en 5 worden een aantal experimenten besproken waarin is onderzocht hoe luisteraars de door prosodie veroorzaakte variatie in het spraaksignaal verwerken bij het fortis-lenis-categoriseren van waargenomen spraakklanken tijdens de herkenning van gesproken woorden, of bij het activeren van woordvormen in het lexicale geheugen.

Prosodische structuur en foneemcontrasten in spraakproductie

De experimenten die besproken worden in Hoofdstuk 2 en 3 gaan over het verband tussen prosodische versterking en het fonologische fortis-lenis-contrast in spraakproductie. Hoofdstuk 2 beschrijft een experiment waarin de standaard versterking van Duitse plofklanken aan het begin van een prosodisch domein is onderzocht. Sprekers lazen verscheidene zinnen waarin de fortis plofklanken /p,t,k/ en de lenis plofklanken /b,d,g/ voorkwamen aan het begin van woorden die minimaal van elkaar verschilden, zoals backen [bakən] ("bakken") en packen [pakən] ("pakken"). De zinnen waren zo gemaakt dat de sterkte van de prosodische grens voorafgaand aan het doelwoord varieerde. Een post hoc prosodische analyse van de opnamen bevestigt dat de spreker deze prosodische grenzen inderdaad onderscheidt. Akoestische metingen van verscheidene bekende kenmerken die een aanwijzing vormen voor het fortis-lenis-onderscheid wijzen op een ingewikkeld verband tussen prosodische versterking en het fonologische contrast. In tegenstelling tot eerder onderzoek in andere talen (bijvoorbeeld Koreaans, Cho en Jun, 2000; Nederlands, Cho en McQueen, 2005) werd gevonden dat niet alle kenmerken door prosodie beïnvloed werden, zoals voorspeld wordt door de twee theoretische verklaringen die tot nu zijn opgesteld voor prosodische versterking: de Uniform Strengthening account en de Feature Enhancement account. Uniform Strengthening zou er toe leiden dat alle plofklanken in toenemende mate fortis worden in sterke prosodische posities, terwijl Feature Enhancement voor fortis plofklanken hetzelfde voorspelt, maar ook voorspelt dat lenis plofklanken in toenemende mate lenis worden. De resultaten waren echter met geen van beide theorieën in overeenstemming. In plaats daarvan wijzen de resultaten op een principe waarin contrast vóór prosodie gaat (zie ook Cole et al., 2007): prosodische versterking heeft de grootste invloed op akoestische eigenschappen die niet fungeren als primaire fonetische kenmerken van fonemische contrasten.

Een vergelijkbaar resultaat werd gevonden in het onderzoek naar de progressieve assimilatie van fricatieven in Hoofdstuk 3. In de inventaris van Duitse fricatieven bestaat een fortis-lenis-contrast tussen /f/ en /v/, en tussen /s/ en /z/. Echter, met uitzondering van enkele woorden die van oorsprong Engels zijn, beginnen Duitse woorden niet met de klank /s/, terwijl er veel Duitse woorden zijn die met de klank /z/ beginnen. De lenis fricatieven /v/ en /z/ worden geassimileerd aan voorafgaande stemloze obstruenten, dat wil zeggen, ze worden stemloos. Dit proces vindt ook plaats wanneer zich tussen de klanken een woordgrens bevindt. Een akoestische analyse heeft aangetoond dat de stemassimilatie gradueel van aard is, en bepaald wordt door de sterkte van de prosodische grens tussen de obstruent en de fricatief. Over prosodische woordgrenzen heen was er minder trilling van de stembanden, oftewel meer assimilatie, dan over de grenzen van prosodische frasen. Dit resultaat komt overeen met bestaand onderzoek dat aantoont dat prosodische versterking leidt tot een grotere weerstand tegen coarticulatie (Cho, 2004). Verder werd gevonden dat /z/ over het algemeen meer verstemloosd wordt dan /v/, een bevinding die verklaard kan worden doordat /v/ aan het begin van een woord de fortis tegenhanger /f/ heeft, terwijl verstemlozing van /z/ niet leidt tot verwarring over de klank, omdat er in het Duits geen woorden bestaan die met /s/ beginnen. De duur van de fricatieven werd ook beïnvloed door prosodische structuur. Na frasegrenzen werd een langere klankduur gevonden dan na woordgrenzen. Met andere woorden: de grotere mate van stemassimilatie die lenis fricatieven makkelijker te verwarren zou maken met hun fortis tegenhangers bij woordgrenzen wordt tegengewerkt door hun korte duur, die eigenlijk te kort is om de verstemloosde fricatieven fortis-achtig te laten lijken. Samenvattend kan worden gezegd dat er geen eenduidige invloed van prosodische effecten op de verschillende kenmerken van het fonologische contrast gevonden is.

Het onderzoek naar de productie van plofklanken en de productie van fricatieven toont onafhankelijke en vaak aanvullende invloeden van prosodie op de verschillende akoestische kenmerken van de fortis-lenis-contrasten aan, zodat deze contrasten altijd worden behouden. Sommige van deze bevindingen verlenen steun aan het recentelijk door Cole et al. (2007) voorgestelde perspectief dat prosodische versterking met name invloed heeft op akoestische eigenschappen van spraakklanken die niet in de eerste plaats kenmerken van fonologische contrasten zijn.

Prosodische structuur en compensatie voor assimilatie in spraakwaarneming

Het productieonderzoek dat in de voorafgaande sectie is samengevat wijst erop dat ondanks de systematische prosodische effecten op sommige akoestische eigenschappen van spraakklanken, prosodische versterking nooit leidt tot lexicale ambiguïteit. Daarom kun je je afvragen of luisteraars deze systematische variatie in het akoestisch signaal gebruiken bij woordherkenning. Hoofdstuk 4 doet verslag van drie foneemidentificatie-experimenten die ontworpen zijn om verder onderzoek te doen naar stemassimilatie van fricatieven in spraakproductie, zoals beschreven in Hoofdstuk 3.

Luisteraars hoorden testwoorden waarvan de beginklanken vervangen waren door spraakklanken die varieerden van volledig stemhebbende tot deels stemhebbende tot volledige stemloze fricatieven. De taak van de luisteraar was om deze fricatieven te identificeren als fortis of lenis fricatieven, in verschillende segmentele contexten (d.w.z., die wel of niet assimilatie toestaan) en verschillende prosodische contexten (na een woordgrens of na een frasegrens). De resultaten tonen aan dat luisteraars compenseren voor stemassimilatie, net zoals ze dat doen voor andere vormen van assimilatie (bijvoorbeeld, Coenen et al., 2001; Gow, 2001; Mitterer en Blomert, 2003). In de experimenten die worden besproken in Hoofdstuk 4 beoordeelden luisteraars deels verstemloosde fricatieven vaker als lenis in de assimilatiecontext dan in de niet-assimilatiecontext. Voorts leken luisteraars in hun compensatie gevoelig voor prosodische structuur: in een assimilatiecontext werden stemloze fricatieven vaker als lenis beoordeeld na woordgrenzen dan na frasegrenzen. Tenslotte was er ook nog een lexicale beperking op de effecten van prosodische structuur. Een prosodisch effect van compensatie voor assimilatie werd namelijk wel gevonden voor het lexicaal functionele /f-v/ contrast, maar niet voor het niet-functionele /s-z/ contrast. Deze bevindingen wijzen erop dat prosodische analyse van gesproken taal een rol speelt bij het oplossen van lexicale ambiguïteit ten gevolge van assimilatie.

De experimenten die besproken werden in Hoofdstuk 4 kennen twee beperkingen. Allereerst wordt foneemcategorisatie slechts beïnvloed door late stadia in het verwerken van gesproken taal. Ten tweede werden, door het groot aantal stimuli en het herhaaldelijk aanbieden van deze stimuli, de testklanken voor ieder fricatiefcontrast slechts voor één paar van minimaal verschillende woorden getest. Daarom werd het in Hoofdstuk 5 beschreven aanvullende onderzoek naar perceptuele compensatie voor stemassimilatie en de beïnvloeding van dit proces door prosodie ondernomen, met behulp van een primingmethode. Deze methode houdt in dat luisteraars een lexicale-decisietaak uitvoeren op een visueel gepresenteerd woord, direct nadat zij een auditieve stimulus (de prime) hebben gehoord. Dit doen zij door op een knop te drukken waarmee ze aangeven of het woord bestaat of niet. Wanneer het woord dat de proefpersoon ziet overeenkomt met het woord dat zij vlak daarvoor gehoord heeft (bijvoorbeeld, de luisteraar hoort en ziet Wälder [vɛldɐ] 'bossen'), kan ze meestal sneller beslissen dat het woord dat ze ziet bestaat. Dit proces duurt langer wanneer de auditieve stimulus een lexicale concurrent heeft geactiveerd, bijvoorbeeld als de luisteraar Wälder [vɛldɐ] 'bossen' hoort, maar Felder [fɛldɐ] 'velden' ziet. Door een woord van een minimaal verschillend paar aan te bieden als targetwoord (in onze experimenten was dit het woord dat met een /f/ begon) kan vastgesteld worden of een stimulus die vlak daarvoor door de proefpersoon is gehoord, waargenomen is als verschillend van of als identiek aan de target. Hierdoor kan dan worden bepaald of de primingeffecten beïnvloed worden door de segmentele of prosodische context waarin de prime is gepresenteerd. De resultaten van dit experiment bevestigen, algemeen genomen, de resultaten van de foneemcategorisatie-experimenten. Na prosodische frasegrenzen compenseerden luisteraars stemassimilatie in de segmentele omgeving die assimilatie mogelijk maakt. Stemloze fricatieven vertraagden reacties op woorden die met een /f/ begonnen in een assimilatiecontext, wat aangeeft dat ze werden waargenomen als goede voorbeelden van concurrerende woorden die met een /v/ beginnen. In een context waarin stemassimilatie niet op kon treden, werden deze klanken waargenomen als ambigue klanken tussen /v/ en /f/. Dit effect werd niet gevonden wanneer de klanken volgden op woordgrenzen. Een mogelijke verklaring hiervoor is dat de duur van de fricatieven werd bijgesteld in de nabijheid van de prosodische grens, en dat de stimuli daardoor misschien te kort waren om het context-effect te laten ontstaan. Dit verschil tussen de prosodische condities is op zich opmerkelijk, omdat het de hypothese bevestigt dat prosodische structuur compensatie voor assimilatie beïnvloedt.

Toekomstig onderzoek

Hoewel de resultaten uit Hoofdstuk 4 duidelijk aangeven dat de mate waarin luisteraars compenseren voor assimilatie wordt beïnvloed door de mate waarin prosodie de assimilatie bepaalt, werd dit effect alleen waargenomen in het geval van een /f/-/v/ contrast, en niet voor de niet-onderscheidende /s/-/z/ tegenstelling. Het blijft een open vraag of dit verschil verklaard kan worden door een fonologische of een lexicale beperking. Voor het zowel fonologisch legale als lexicaal bestaande onderscheid tussen Wälder [vɛldɐ] ‘bossen’ en Felder [fɛldɐ] ‘velden’ waren luisteraars gevoelig voor de prosodische variatie in de mate van verstemlozing. Als de fonologie hiervoor verantwoordelijk is, zouden we hetzelfde prosodische effect voorspellen wanneer de targets bestaan uit het fonologisch legale /f/-/v/ onderscheid, zelfs wanneer dit getest wordt bij een woord versus nonwoord, bijvoorbeeld Wasser [vasɐ] ‘water’ en *[fasɐ]. Echter, wanneer de effecten afhangen van de lexicale status van de eindpunten van het continuüm van stimuli, zouden de reacties van proefpersonen van dezelfde aard moeten zijn als die in het Senken [zɛŋkən] ‘dalen, valleien’ versus *[sɛŋkən] experiment, waarin geen effect van prosodie werd gevonden. Een dergelijk onderzoek zou meer inzicht geven over de vraag waar in het spraakherkenningssysteem compensatieprocessen plaatsvinden (op het pre-lexicaal fonologische niveau of op het lexicale niveau).

Algemene implicaties

Uit het onderzoek dat in dit proefschrift wordt besproken kan afgeleid worden dat prosodie en het segmentale niveau vervlochten, maar onafhankelijke systemen zijn. Uit de productie-onderzoeken bleek dat prosodie de akoestische details van spraakklanken op een systematische manier beïnvloedt, maar dat dit proces zich beperkt tot akoestische dimensies die niet gebruikt worden om het fonologisch onderscheid te maken. Echter, onder zorgvuldig opgezette en gecontroleerde experimentele omstandigheden kon aangetoond worden dat luisteraars opmerkelijk gevoelig zijn voor alle systematische variatie in het spraaksignaal, en alle daarin beschikbare informatie gebruiken bij het uitvoeren van een spraakherkenningstaak.

Dit proefschrift draagt bij aan voorgaand onderzoek naar de rol van door prosodie bepaald fonetisch detail in spraakproductie (bv., Cho en Keating, 2001, 2007; Fougeron en Keating, 1997; Fougeron, 2001; Keating et al., 2003; Tabain, 2003a, 2003b) en woordsegmentatie in spraakperceptie (bv., Salverda et al. 2003; Shatzman en McQueen, 2006; Cho et al., 2007) en breidt dit onderzoek uit tot de fonemische identificatie van spraakklanken.

Referenties

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

Cho, T. and Jun, S-A. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistic Society* **36,** 31–44.

Cho, T. and Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* **29**, 155-190.

Cho, T. and McQueen, J.M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* **31**, 121-157.

Cho, T., McQueen, J. M., and Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* **35**, 210 - 243.

Coenen, E., Zwitserlood, P. and Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* **16**, 535 - 564.

Cole, J., Kim, H., Choi, H. and Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* **35**, 180-209.

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

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

Keating, P. A., Cho, T., Fougeron, C. and Hsu, C.-S. (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.

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

Salverda, A., Dahan, D. and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* **90**, 51-89

Shatzman, K. B. and McQueen, J. M. (2006) Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics* **68**, 1-16

Tabain, M. (2003a). Effects of prosodic boundary on /aC/ sequences: acoustic results. *Journal of the Acoustical Society of America* **113**, 516-530.

Tabain, M. (2003b). Effects of prosodic boundary on /aC/ sequences: articulatory results. *Journal of the Acoustical Society of America* **113**, 2834-2849.

Curriculum vitae

Claudia Kuzla was born on March 28, 1973 in Neustadt/Holstein, Germany. Supported by a scholarship from the *Studienstiftung des Deutschen Volkes,* she studied General Linguistics, Phonetics, and Philosophy at the Christian-Albrechts-Universität zu Kiel, Germany. A grant from the *Deutscher Akademischer Auslandsdienst* (DAAD) allowed her to spend a semester at the Phonetics Department of Stockholm University, Sweden in 1996. After a period of maternity leave, she received her M.A. in General Linguistics from Kiel University in December 2001. In 2002, she was awarded a PhD scholarship from the *Max-Planck-Gesellschaft zur Förderung der Wissenschaften* and joined the Comprehension Group at the Max-Planck-Institut für Psycholinguistik in Nijmegen, The Netherlands, where the research described in this thesis was carried out. Currently, she holds a postdoctoral fellow position at the Institut für Phonetik und Sprachverarbeitung (IPS) der Ludwig-Maximilians-Universität München, Germany.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*
27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*
35. Retrieving semantic and syntactic word properties. *Oliver Müller*
36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*
37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*
38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*
42. The acquisition of auditory categories. *Martijn Goudbeek*
43. Affix reduction in spoken Dutch. *Mark Pluymaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*
45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*

NOTES

NOTES