Production and Perception of Word Boundary Markers in German Speech

Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie (Dr.phil.)

vorgelegt von

Eisenhuth, Heike

an der

Universität Konstanz

Geisteswissenschaftliche Sektion

Fachbereich Sprachwissenschaft

Konstanz, 2015

Referentin: Prof. Dr. Bettina Braun
Referent: Prof. Dr. Henning Reetz
Acknowledgements

This dissertation would not have been possible without the help of a great many of people. First of all, I want to thank my supervisors Bettina Braun and Henning Reetz who guided and inspired me and patiently answered all my questions. This thesis has been written parallel to my work at the University of Applied Sciences Konstanz and later at the University of Music and Theatre Rostock. My superiors Werner Hofacker and Frank Ivemeyer agreed to very flexible working hours and thus enabled me to pursue my research as effectively as possible. I want to thank them for their courtesy. I am also grateful to so many students, teachers, and colleagues from administration of both universities and to my family and friends for giving me their time and taking part in my experiments. My brother Thomas Eisenhuth helped me with the organisation of my recordings. His support was much appreciated. Special thanks to Karsten Robel, Daniela Trunk, Pascal Zurek as well as to Hartmut Möller and the PhD students of the University of Music and Theatre Rostock who gave me valuable input for my experimental designs. I am also much obliged to my friends and family members who proofread this dissertation: Kirstin Blös, Karsten Robel, Zuzana Robel-Högerova, Lesley Taylor and Margaret Wilson. Very special thanks I owe to Silvio Robel for his constant support. Not only has he written several programmes which I used for my experiments and guided me through LaTex, but also encouraged me to go on in times when working on my dissertation was difficult. His critical assessment of my work and his suggestions helped me tremendously. Finally, I want to thank my parents who always believed in me.
Contents

Glossary

1. Introduction
   1.1. Word production .............................................. 5
   1.2. Word perception models and processes ......................... 11
   1.3. Outline of the thesis ........................................ 21

2. Theoretical background: Word boundary markers in German speech 22
   2.1. Glottal stops/creaky voice .................................... 22
       2.1.1. Production .............................................. 25
       2.1.2. Perception .............................................. 27
   2.2. Stop allophones .............................................. 27
       2.2.1. Production .............................................. 28
       2.2.2. Perception .............................................. 29
   2.3. L-allophones ................................................. 30
       2.3.1. Production .............................................. 31
       2.3.2. Perception .............................................. 32
   2.4. Silent intervals .............................................. 33
       2.4.1. Production .............................................. 34
       2.4.2. Perception .............................................. 35
   2.5. Stress ......................................................... 36
       2.5.1. Production .............................................. 38
       2.5.2. Perception .............................................. 40
   2.6. Comparative vowel duration .................................... 43
       2.6.1. Production .............................................. 43
       2.6.2. Perception .............................................. 44
   2.7. Geminates .................................................... 45
       2.7.1. Production .............................................. 47
### 4.3. Experiment 2

- **4.3.1. New Recording** .................................................. 99
- **4.3.2. Methods** .......................................................... 100
  - **4.3.2.1. Participants** ............................................... 100
  - **4.3.2.2. Materials** .................................................. 101
  - **4.3.2.3. Procedure** ............................................... 105
- **4.3.3. Hypotheses and research questions** ......................... 106
- **4.3.4. Results** ......................................................... 109
- **4.3.5. Discussion** ..................................................... 114

### 5. Summary of the results, evaluation of experimental conditions, outlook

- **5.1. Experimental results** ............................................. 120
- **5.2. Evaluation of the experimental conditions in the perception tests** 123
- **5.3. Future Research** ................................................ 124
- **5.4. Zusammenfassung** ............................................... 125

### A. Appendix

- **A.1. Reading Texts** .................................................... 147
  - **A.1.1. Production experiment, text 1** .......................... 147
  - **A.1.2. Production experiment, text 2** .......................... 148
  - **A.1.3. Production experiment, text 3** .......................... 148
  - **A.1.4. Production experiment, reading text for the second singleton-geminate analysis** .................................................. 151
  - **A.1.5. Perception experiment, additional recording** ........ 152
- **A.2. List of Experimental Materials** ................................ 157
  - **A.2.1. List of Targets** .............................................. 157
    - **A.2.1.1. Perception experiment 1** ............................. 157
    - **A.2.1.2. Perception experiment 2** ............................. 159
  - **A.2.2. Singleton-Geminate Targets** ............................... 162
    - **A.2.2.1. Test 1 (uncontrolled)** ............................... 162
    - **A.2.2.2. Test 2 (controlled)** ................................. 163
## Glossary

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>/e/</td>
<td>phoneme</td>
</tr>
<tr>
<td>&lt;e&gt;</td>
<td>morpheme</td>
</tr>
<tr>
<td>[e]</td>
<td>phonetic realisation of a phoneme</td>
</tr>
<tr>
<td>[merger context]</td>
<td>Brackets show the context out of which the merger has been taken</td>
</tr>
<tr>
<td>acc.</td>
<td>accusative</td>
</tr>
<tr>
<td>dat.</td>
<td>dative</td>
</tr>
<tr>
<td>ERP</td>
<td>Event related potential</td>
</tr>
<tr>
<td>fem.</td>
<td>feminine</td>
</tr>
<tr>
<td>FUL</td>
<td>Featurally Underspecified Lexicon</td>
</tr>
<tr>
<td>IP</td>
<td>intonational phrase</td>
</tr>
<tr>
<td>L2</td>
<td>second language</td>
</tr>
<tr>
<td>masc.</td>
<td>masculine</td>
</tr>
<tr>
<td>mus.</td>
<td>musical</td>
</tr>
<tr>
<td>neut.</td>
<td>neuter</td>
</tr>
<tr>
<td>nom.</td>
<td>nominative</td>
</tr>
<tr>
<td>p.</td>
<td>person</td>
</tr>
<tr>
<td>PBS</td>
<td>prosodic boundary strength</td>
</tr>
<tr>
<td>pl.</td>
<td>plural</td>
</tr>
<tr>
<td>SG</td>
<td>Standard German</td>
</tr>
<tr>
<td>sg.</td>
<td>singular</td>
</tr>
<tr>
<td>SW</td>
<td>strong-weak</td>
</tr>
<tr>
<td>TOT</td>
<td>tip of the tongue</td>
</tr>
<tr>
<td>wb</td>
<td>word boundary</td>
</tr>
<tr>
<td>WS</td>
<td>weak-strong</td>
</tr>
</tbody>
</table>
1. Introduction

This work investigates the acoustic manifestation of word boundary markers in German speech production and their impact for perception. While in the written form words are separated by white spaces, spoken phrases consist of an uninterrupted sequence of words (Ingram, 2007). The signal shows an ongoing audio stream; sometimes there are pauses, but these pauses do not necessarily coincide with word boundaries. Speech involves a wide range of dynamic processes, which makes the listener’s decision about where a word starts and where it ends even more complicated, because word internal and across word processes can alter the acoustic shape of a word. Some German examples shall illustrate this. In the remainder of this thesis, the translation of examples follows after a slash.

1. Words can be cliticised which might lead to re-syllabification (konnte ich/could I [kʰɔntə ʔɪc] → [kʰɔntɪtɪc]).

2. Phonemes can alter their place of articulation (Alpen/Alps → [αlp]).

3. Phonemes can be reduced or fully eliminated (mit dem/with the, dat. → [mɪpm], [mɪm], [mm] or [mɪtm], example taken from Kohler, 1998, p. 22).

However, general linguistic knowledge about phonetic and phonological rules in connection with the ability to process syntactic, morphological, and semantic information enable listeners to distinguish one word from another as they perceive the speech stream. Listeners also analyse paradigmatic and syntagmatic relations between phonemes, syllables and words. For example, the word Kleid/dress in the phrase das grüne Kleid/the green dress has a paradigmatic relationship with every other item of the language that can be exchanged for it: Rock/skirt, Hügel/hill, etc. The article das/the (neut.) has a paradigmatic relationship with e.g. die/the (fem.), unser/our, ein/a; adjective grüne/green with schöne/beautiful (fem.), alter/old (masc.), gebrauchtes/used (neut.) and so

---

1A brief discussion regarding the domain of the word and what definition applies to this thesis is given later in this chapter.
on. Syntagmatic relations between items arise from the construction. For example, the sequence “feminine article” + “adjective” is usually followed by a feminine noun.

Paradigmatic and syntagmatic relations are also interpretable from the phonological perspective. A set of phonemes, e.g. /b/, /t/, /s/, has a paradigmatic relation in the context “au” - Bau/construction site, Tau/dew, Sau/sod. Syntagmatic relations exist with regard to the ordering of phonemes in the word. Every language has rules concerning legitimate and illegitimate phoneme sequences with regard to their position in a syllable. In German, for instance, the consonant cluster /mt/ does not appear word-initially, but it can appear word-medially or -finally (Amter/offices, kramt/rummages).

In this thesis, we focus on the word level to examine the role of wb markers. Wb markers are phonetic cues in the audio stream and signal the boundaries of words. A silent interval between two spoken words is an example for a wb marker. We will further outline the functionality of wb markers using the following sequence, which represents an ongoing, uninterrupted speech stream and is thus presented without white space between the words.

diechefinderenreporter...

Two interpretations are possible:

1. Die Chefin, deren Reporter/The boss (fem.) whose reporter
2. Die Chefin, deren Report er/The boss (fem.) whose report he

The presence or absence of a wb marker between Report and er supports disambiguation. In our case, the presence of a wb marker between /t/ and the following vowel would work in favour of the second interpretation. Possible wb markers for the given example are:

1. a glottal stop before er → [ʔeːy]
2. er carrying an initial creaky voice segment → [eːy] - possibly in conjunction with a glottal stop
3. a silent interval between Report and er
4. the vowel quality and quantity of /e/ in the pronoun er → [ʔeːy] compared to the morpheme <-er> → [i]
1. Introduction

The presented example already named some markers that we investigated. Glottal stops, creaky voice, and silent intervals are three of the seven markers that were analysed. The other four are position-specific allophones (exemplarily, stop and /l/-allophones were examined), stress, comparative vowel duration, and contact geminates (exemplarily, nasal contact geminates and nasal singletons were contrasted) - see Table 1.1 for an overview. These markers were chosen because we expected to get new insights into segmentation strategies during the production and perception of speech. All markers mentioned above have been the object of intensive prior research. Their appearances and boundary marking potentials are well documented, especially for German, the language under discussion in this work. Comparative vowel duration as proposed by Rietveld (1980) has mostly been examined for French. Rietveld found that in French, the different vowel durations in two phonemically identical but syntactically different phrases supported disambiguation to some effect.

This thesis investigates the aforementioned group of markers systematically for German speech. As a novelty, we evaluate the chosen markers’ individual boundary marking strength. Relations between production and perception data are shown as well as interactions of markers. Experiments were conducted in the majority with participants who had no or very little linguistic background.

For our own analysis, the creation of a special speech corpus with words carrying the markers of interest was a central point. The corpus became the basis for the perception experiment described in Chapter 3. Markers were examined with regard to their production specifics. Here, only unpredictable markers were analysed. The term “unpredictable” refers to the fact that it is uncertain whether speakers use them or not.

A two-part perception experiment (chapter 4) tested the functional importance of markers in the process of juncture perception. The focus was on syntagmatic recognition processes. Assuming that markers have a different impact on the perception system, experimental findings are expected to lead to a ranking list of markers according to their importance for speech comprehension.

When markers are discussed, the domain of the word has to be defined first. Generally, there is no standard definition available for what a word is. The topic has been the subject of an ongoing controversy, which is briefly summarised here. The word can

---

2 The term “contact geminate” is used for a segment that arises when one word ends and the following one starts with the same phoneme, for example: *einem Mann/a man, dat.*
### 1. Introduction

<table>
<thead>
<tr>
<th>Word boundary marker</th>
<th>Sequence containing the word boundary marker</th>
<th>Sequence without a word boundary marker</th>
<th>Marker specifics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal stops/creaky voice</td>
<td>mag-eR/‘likes he’</td>
<td>mager/skinny</td>
<td>Often found in words with initial vowels.</td>
</tr>
</tbody>
</table>

#### Allophones

- **Stop allophones**
  - [erst]e’s Kalb/first calf
    - (Also note that final /b/ is devoiced.)
  - Skalp/scalp
    - Stop allophones may differ in aspiration. Word-initially (Kalb), the stop is aspirated, in a cluster with /s/, however, the stop is not aspirated.

- **/l/-allophones**
  - Pappe l[liegt]/cardboard lies
  - Pappel/poplar
    - The neighbourhood of /l/ determines allophonic variation.

#### Silent intervals

- Koda#k[l]ingt]/coda sounds
  - (The hash shows a potential place for a silent interval).
  - Kodak/name of a company
  - In general, silent intervals have an important function in speech segmentation. In the sequence Koda#k, a silent interval after /a/ resulting from a stop closure coincides with a wb.

#### Stress

- [dass Wind und] Wasser wirtschaft[l]ich sind]/that wind and water are profitable
  - Wasserwirtschaft/water economics
  - Stress influences the activation of word candidates in the mental lexicon.

#### Contact geminates (exemplarily, nasals were chosen)

- /m/ in einem Mann/one man, dat.
  - /m/ in einem Tag/one day, dat.
  - Contact geminates and singleton segments contrast in duration.

#### Comparative vowel duration

In French, the disambiguation of phonemically identical, but syntactically different sequences is possible. The compared sequences showed opposing duration patterns for the vowels (short-long vs. long-short). This was also tested for German, e.g. duration of diphthong /au/ and vowel /a/ in [mit den] Augenmaß/measured by the eye as compared to duration of diphthong /au/ and vowel /a/ in Augenmaß/sense of proportion.

**Table 1.1.** The examined word boundary markers and their specifics.
be seen from various perspectives. First, there is the **orthographical word** of the written language. Orthographic words are separated by white spaces. The **grammatical word** consists of elements which always occur together and in a fixed order with conventionalised coherence and meaning (Dixon and Aikhenvald, 2002). An example for a grammatical word is *Peter der Große/Peter the Great*. The **phonological word** may fall together with the grammatical word, but it may also comprise two or more grammatical words (Dixon and Aikhenvald, 2002). This might be the case, when words are cliticised in casual speech, as mentioned before. A **lexeme** refers to a lexical unit that carries meaning. In German, the lexeme usually refers to a word or the word stem (Kühn, 1994). *Bauen/to build* represents the lexeme for forms like *baut/builds 3. p. sg.*, *baute/built 3rd p. sg. past tense*, *gebaut/built past participle*, etc. Finally, there is the **morpheme** as the smallest meaningful unit of a language. It participates as a single unit or in connection with other morphemes in the construction of a word form (Simmler, 1998, p. 65). The word *Stiere/bulls*, for instance, consists of two morphemes: *<Stier>* and *<e>* where *<Stier>* represents the lexical morpheme as the smallest meaningful unit, and *<-e>* is a suffix allomorph that signals “plural”. The *wb* markers that will be discussed in this thesis refer to lexemes.

The next section explains the mechanisms involved in word production and how phonological information is encoded through the application of *wb* markers.

### 1.1. Word production

In the following, the process of word production is seen from two perspectives. Firstly, we look at how the word is produced as an individual unit. Secondly, the word will be seen as part of a larger sequence. This is of special interest, since all targets used for our experimental part were embedded in sentences. In running speech, the shape of the word can alter as explained above. Boundaries become blurred, the beginnings and ends of words interact with left and right neighbours.

A word is produced through the generation and alignment of syllables, which are assumed to be the core units in the articulation process (Levelt, 1993). This assumption implies that every adult has an inventory of syllables stored in the mental lexicon (ibid., p. 318). A word may consist of one syllable. Larger words are constructed through the alignment of syllables according to a phonetic plan. As part of the phonetic plan, the speaker
1. Introduction

has to be aware of possible word beginnings and endings. This includes, for instance, knowledge about legitimate word-initial and -final consonant clusters, devoicing of voiced final stops (as in German), resyllabification etc.

A fruitful way to study phonetic plans for words is the analysis of speech errors and the so-called tip of the tongue (TOT) phenomenon, a state “where the transition from lemma to sound form is hampered” (Levelt, 1993, p. 320). A word lies on the tip of the tongue, but cannot yet be produced.

Brown and McNeill (1966) constructed an experiment in which they brought test persons into the TOT state in order to get insights into the word production process. Their experiment showed how much lexical-form information is available, before a word is articulated. Participants received dictionary definitions of low-frequency words, for instance, “a navigational instrument used in measuring angular distances, especially the altitude of sun, moon, and stars at sea” and the task was to tell which object was described. The correct answer was ‘sextant.’ Some test persons knew the answer immediately, some had no idea at all, some produced words that were similar in meaning to the target (e.g. astrolabe, compass). The results Brown and McNeill were especially interested in were the responses of participants which had come into a TOT state. These participants were on the verge of producing the correct answer. They were asked for more information: what is the initial letter, how many syllables does the word have, which words are in their mind, etc. Test persons showed a tendency to guess /s/ as the initial phoneme and ‘2’ as the number of syllables, and that primary stress was on the first syllable. Among the words that had come to their minds was ‘secant.’ The experimental results demonstrate how much information was present before the word had even been uttered: the first phoneme or cluster and possibly also the stress placement.

Stress distribution is an important factor for word segmentation, which leads us to speech errors connected with stress. They also provide a valuable source for the understanding of information processing prior to the actual articulation of a word. The issue of word stress perception and consequences of mis-stressing will be followed-up in the experimental part of this thesis. Cutler (1980) presented a large speech corpus with words containing stress errors. One example is the sentence You’re in a real advantageous position. The speaker had started to utter the adjective advantageous with stress on the second syllable as it would be the case for the related word advantage. One conclusion is that in the moment of the morphological/metrical spellout (as Levelt calls it), the target
advantageous, as well as the related word ad\textit{vantage}, were activated\footnote{According to Levelt (1993), the moment of the morphological/metrical spellout is when a speaker retrieves the morphemes and metrical structure of a word by using lemmas stored in the mental lexicon.}. The given example shows how stress placement (word production) and word segmentation (perception) are connected. While the originally chosen word represents a legitimate sequence in general, it violated syntactic demands. The word’s stress pattern was associated with a noun which in our example was incompatible with the previously chosen verb. In order to avoid an ungrammatical utterance, the speaker rejected the initially chosen word and selected a new one. A question worth considering is whether a listener on the other hand tolerates an incorrectly stressed word like ad\textit{vantageous} when it is perceived as an isolated target. The sequence is meaningful, and the wrong stress pattern might be forgiven. Cutler and Clifton (1984), Domahs et al. (2008) and others have shown that misstressed words might be segmented like the correctly stressed word under certain conditions.

Bringing the topic of stress back to the field of word production, the research of Liberman and Prince (1977) regarding stress shift highlights an interesting aspect. Being part of a larger unit, the authors stated, the word can receive a stress pattern that differs from the canonical citation form. A reason may be that the word subordinates to the rhythmic organisation of an utterance. If words would simply line up with their canonical stress patterns, stressed syllables may sooner or later directly follow each other and disrupt the rhythmic alternation of the utterance which would lead to a stress clash between two adjacent syllables. Selkirk (1986) introduced the rule of Beat Movement which resolves such a stress clash. Levelt (1993) used the example \textit{sixteen dollars} to illustrate the rule: With a preprocessing of \textit{dollars} as the upcoming word, the stress pattern for \textit{sixteen} changes and the clash is prevented (see Figure 1.1). However, Levelt (1993) also pointed out that there are phrasal restrictions on beat movement. Should the intended phrase be \textit{Dimes I have sixteen, dollars just one}, no beat movement takes place. The word \textit{sixteen} is followed by a phrase boundary which prevents it.

Another method to structure an utterance is the insertion of so-called silent beats (Selkirk, 1986) which can take several forms - the phrase-final lengthening of a syllable, for instance, or the insertion of a silent interval. Final lengthening of a syllable as well as silent intervals have been examined in greater detail in the upcoming production and in the perception experiments. The focus was on their places of occurrence and on their durations. One major question we pursued was whether deliberately inserted silent
1. Introduction

Figure 1.1. Example for a beat movement assuming that “six” has secondary stress. The metrical grid shows the weight of a syllable as the sum of beats aligned on different layers. Left: stress clash pattern, right: stress pattern after beat movement.

intervals were distinguishable from silent periods that characterise a sound (e.g. stops) or not. For instance, the sequence Koda#k - the hash shows the place of the silent interval - may be one word, namely Kodak/name of a company, and the silent interval between /a/ and /k/ is the stop closure period. On the other hand, the sequence might represent the word Koda/coda plus the first phoneme of the following word klingt/sounds, verb. We assume that the silent interval between /a/ and /k/ in the latter sequence might be longer because the stop closure and the fall together.

In addition to silent intervals, phrase-final lengthening is another break option mentioned by Selkirk (1986). This topic will be resumed when we discuss contact geminates in German speech, see section 2.7.

Another point that is relevant for our experiments is lexical bias. We have mentioned earlier in this section in connection with the phonetic plan that the adjective advantageous was activated along with the related word advantage. Note that the incorrectly activated word was meaningful. This may be explained in terms of lexical bias (Levelt, 1993). Sub-lexical errors, for instance, produce words rather than non-word strings. Baars, Motley, and Mackay (1975) presented a solid foundation for this assumption. They created an error elicitation experiment in which they asked test persons to read a list of word pairs, e.g. darn bore, dart board. In order to prompt speech errors, three bias pairs (e.g. ball dome) had to be uttered before the target pair. Baars, Motley, and Mackay (1975) found that the artificially elicited slips of the tongue created more real words (barn door) than nonsense words (bart doard). Lexical bias is also an interesting topic from the perception
point of view and will be re-addressed in the next section which introduces perception processes.

Apart from speech errors concerning the morphological/metrical spellout, we often find speech errors at the phonetic spellout level (Levelt, 1993). The production studies introduced in the following highlight the importance of the initial syllable, which played an important role in our own experiments, too, because initial syllables are often preceded by or carry [wh] markers. Fromkin (1973) studied a variety of speech errors and found a large group in which the initial phonemes were exchanged, e.g. *heft lemisphere* (*left hemisphere*). In this example, /h/ was apparently more strongly activated than /l/. After the wrong phoneme (/h/) was used in one word onset, /l/ was still available and used for the next fitting location, namely the next following word onset. Presumably the exchange was possible, because both segments were of the same category (onsets). Shattuck-Hufnagel and Klatt (1979) followed this line of argumentation. They stated that target intrusions happen according to the principle of the “Unit-similarity constraint”. According to this principle, elements involved in a speech error are of the same level of representation and category. Shattuck-Hufnagel and Klatt analysed 1620 spontaneous speech errors (*e.g.* *pade mossible* - *made possible*, *Christmas crob* - *Christmas club*, *fost and lound* - *lost and found*). They too found that that word-initial syllables were frequently involved in speech errors. Shattuck-Hufnagel (1985) came to the same conclusion. In a tongue-twister experiment, she prompted participants to produce speech errors. Test persons were given a card with words printed on it. They were instructed to read the card three times before they turned it over and recited the words from memory. Word-initial speech errors occurred 4.5 times more often than speech errors in syllables that were not in initial position. Shattuck-Hufnagel concluded that a sublexical serial ordering process starts at a very early stage in the production planning process, even before some phonetic processes take place.

The literature presented in this section so far has mainly looked at the production of individual words. Now we place the process of word production in the context of connected speech. Note that the targets used in the experimental part of this thesis were not produced as isolated items but originate from connected speech. A specialty of connected speech is that “a word’s stored syllabification is not sacrosanct” (Levelt, 1993, p. 318). Neighbouring words may interact with each other and lead to re-syllabification: *konnte ich/could I* [kʰɔntə ʔɪcf] → [kʰɔntətuc], as mentioned in the introduction. The pronoun
is cliticised, and the verb is reduced from a two-syllable-word to a one-syllable-word. Reduction processes are very common in connected speech. An interesting question is whether word reduction is connected to lexical frequency. The literature shows contradicting data. Zimmerer, Scharinger, and Reetz (2011) could not find such a connection for German. They asked native speakers to produce declination forms of German regular and irregular verbs (e.g. raten/guess: ich rate, du rätst, er rät etc./I guess, you guess, he guesses etc.) and analysed the speech data regarding word-final /t/-deletion. The 2nd person singular forms of irregular verbs have a lower lexical frequency compared to regular verbs. However, /t/-deletion in 2nd person singular suffixes occurred equally often in both verb classes. Pluymaekers, Ernestus, and Baayen (2005), on the other hand, reported that there is evidence for the relationship between lexical frequency and acoustic reduction. In a database analysis for Dutch, frequently used words such as natuurlijk/naturally showed more reduction processes than rarer Dutch words. The maximum reduced form for natuurlijk was [tvk]. Pluymaekers et al. suggested two possible mechanisms to explain why high-frequency words are more prone to reduction processes than low-frequency words.

1. Listeners have stored all reduced forms as redundant information in the mental lexicon. The authors believed this to be unlikely. There are obvious cases like [tvk] for natuurlijk, they argued, but there are many words which allow gradient reduction, and one would have to store all these forms.

2. Words stored in the mental lexicon have a certain resting activation level. Highly frequent words have a high activation level for their respective speech sounds “resulting in quicker preparations of the speech sounds and thus, shorter articulatory durations” (ibid., p. 19).

Reduced word forms were also the topic of a corpus study done by Kohler (2001). He found 68 occurrences of the word eigentlich/normally in the Kiel Corpus of Spontaneous Speech (IPDS, as published on CD-ROM, 1995, 1996, 1997). The lexical citation form [’aŋɪtlɪk] was not present, but various phonetic realisations instead. Not a single form contained a schwa after /g/. Nine forms had dorsal or velar closures for /g/, but eight of these were dorsal adjustments for the nasal /g/. The most extreme form that had been recorded was [’aŋʃ]. There are a great number of other studies that discuss the connection between highly frequent words and reduction processes, e.g. Jurafsky et al. (2001), Bybee (2002) or Shi et al. (2005), to name few.

Apart from reduction, assimilation is another prominent characteristic of (fast) speech.
A speech sound approximates to a neighbouring sound through the adoption of some of the neighbour’s properties. Assimilation is possible in several ways, three examples for German shall illustrate the process:

1. Progressive assimilation of place of articulation (e.g. [pn] → [pm], Alpen/Alps → [ˈalpm], as mentioned in the introduction to this chapter).

2. Progressive assimilation of nasality (e.g. [n] → [ŋ], von Grund auf/from scratch → [fon 'grunt auf])

3. Regressive assimilation of place of articulation (e.g. [nb] → [mb], anbeißen/to bite into something → [ˈambaiʃn])

Zimmerer, Reetz, and Lahiri (2009) have analysed regressive place assimilation in German using the Kiel Corpus of Spontaneous Speech (as published on CD-ROM, IPDS, 1995, 1996, 1997); 7060 possible sequences were examined. They found that 393 instances were in fact assimilated (5.6 per cent), most of all function words (266 cases). Despite being a rare process, we also found place assimilations in our production data, even across the wb, for example in Handwerk lebt/craftsmanship is alive. If the phrase was produced as one phonological word, the resulting utterance was [das 'hantvɐr'ɡleːpt].

In this section, a brief overview of the production of words as individual items and as constituents of an utterance was given. The next section concentrates on the perception process. Since this thesis is concerned with segmentation strategies, the role of wb markers in connection with speech segmentation will be carved out.

1.2. Word perception models and processes

This thesis follows up the question which wb markers support listeners to recognise the beginnings and ends of words correctly. The actual identification of the word is the final stage in a more complex process. Speech is continuous and the boundaries of a lexical unit may be marked differently. Listeners apply segmentation heuristics based on their knowledge of linguistic rules and are thus able to parse the speech stream into words (Cutler, 1992). Wb cues help listeners to perform the segmentation task. Clearly articulated utterances, for instance, are easier to decode by listeners than casual,
inattentive speech. We assume that this is the case, because clear speech usually contains more markers.

The different phases of word perception including segmentation strategies are briefly summarised in the following with reference to Frauenfelder and Tyler (1987), who provided a detailed overview of the topic. They assumed that the first stage in word recognition is the initial contact phase. The speech wave activates lexical entries that are consistent with the signal. For these initial contacts, linguists have proposed different kinds of lexical representations: temporally-defined spectral templates (Klatt, 1980), features (Shipman and Zue, 1982), phonemes (Pisoni and Luce, 1987) or syllables (Mehler and Hayes, 1981), to name few.

The initial contact phase, Frauenfelder and Tyler continued, is followed by a selection phase. The more discriminative the incoming acoustic-phonetic information, the smaller the number of lexical entries which are initially contacted. Lexical entries are assessed against sensory input. Word recognition happens at the end-point of the selection phase. The word recognition point hereby represents the exact moment when the word is recognised (Marslen-Wilson and Tyler, 1980). This point can be reached before the actual word has been completely heard. Warren and Marslen-Wilson (1987) demonstrated this in a perception experiment. Listeners heard monosyllabic word pairs which differed in their final consonant (e.g. *scoop, *scoon, *scoob, *scood, in English.)

In the following, we introduce models of word perception which are important in connection with our own perception experiment.

A prominent word perception model is the Cohort Model by Marslen-Wilson, which in its original version (Marslen-Wilson and Welsh, 1978; Marslen-Wilson and Tyler, 1980) claims that all lexical entries matching the incoming contact representations are activated (a ‘cohort’) and that all of them have an equal status. The original selection of word candidates is based on the acoustic-phonetic properties of the incoming word’s initial segment. Accordingly, the initial cohort contains all words of the respective language that match some beginning portion of the input (Marslen-Wilson and Tyler, 1980). Even
though the first pool of potential words might be large, “at least the system could be sure that the correct word was among them” (Marslen-Wilson and Welsh, 1978, p. 60). As more acoustic-phonetic information becomes available, certain candidates cease to be compatible with the speech input and drop out of the cohort until only one remains. This is the point when the word is recognised.

Later versions of the cohort model (e.g. Marslen-Wilson, 1987; Grosjean and Gee, 1987) pursue a more flexible strategy. Input that fails to match lexical entries is not rejected, but the level of activation declines. Furthermore, the activation of lexical entries depends upon the speech stream’s “goodness of fit” (Frauenfelder and Tyler, 1987, p. 5) and internal specifications such as an item’s frequency of occurrence. More details about the effect of word frequency on word perception is given later in this section. Grosjean and Gee (1987) argued that stretches of the signal that are particularly reliable such as stressed syllables establish the initial contact. A stressed syllable initiates lexical research, weak syllables are located on either side of the stressed syllable. The authors proposed that the listener performs a pattern-recognition task applying phonotactic and morphophonemic rules.

A competing model is the **Trace** model, introduced by Elman and McClelland (1983). It assumes that while hearing a certain sequence, lexical entries are activated as the acoustic signal is unfolding. The basic principle is therefore similar to the cohort model. However, the models differ with respect to where the activation point for a word candidate lies. While the cohort model assumes that each word has a unique contact, Trace as a connectionist model is not restricted to a strict sequential order and allows several contacts. Each activated phoneme initiates a new subset of lexical entries containing this phoneme. Consequently, activation spreads across three levels: the featural level as the lowest one, the phoneme level in the middle, and the word level as the highest level. Candidate pools change constantly as the speech stream continues, rejected candidates receive lower activation and may eventually drop out, potential new ones come in. As a principle, units at each level compete with each other, while units in lower levels activate units in higher levels. For example, for the target word *catalogue*, the candidates *cat* and *catalogue* compete for the initial three phonemes, the candidates *log* and *catalogue* compete for the final three phonemes.

McQueen, Norris, and Cutler (1994) criticised Trace because too many overlapping lexical networks can potentially begin anywhere in the input stream. For a sequence of
50 phonemes, Trace would require 50 complete lexical networks, the authors argued (ibid., p. 622 f.). Norris (1994) tackled this problem when he introduced the **Shortlist model**. Other than Trace, Shortlist works bottom-up. Potential candidates are shortlisted and evaluated in an interactive activation network that functions like the lexical level of Trace. But instead of many overlapping lexical networks, Shortlist assumes only a small competitive network with a defined set of candidates. The syllable in the word onset that gets higher activation, is the stronger competitor. Norris (1994) stated that the bottom-up operation of Shortlist does not restrict the functionality of the model. In his opinion, top-down feedback from the word to the phoneme level is not necessary, because all crucial lexical restraints can be processed on the lexical level itself.

A great challenge for word perception models is that the total recognition point of a word may be detached from the word. We have mentioned before that a word may become unique before its offset. However, sometimes the uniqueness point of a word may not be reached before a part of the following word has been heard. Grosjean (1985) showed that in a gating experiment. He confronted listeners with sequences that ended before the target word had reached its acoustic offset and asked them to finish off the sentence and to rate how confident they were about their choice. The sentences were repeated and each time, more acoustic information was added to the target word. The results of the experiment showed that words were not always recognisable before a portion of the following word had been heard. Shorter words were harder to recognise than longer words. For instance, the recognition of *bun* proved to be difficult for listeners; identification was not possible before the following segment had been heard. This is because the syllable *[bʌn]* represents a word in itself, but also the beginning of many other words, e.g. *bundle, bunny*, etc. (ibid., p. 307). Grosjean (1985) claimed that there has been a “bias in the literature toward immediate, word-by-word recognition” (ibid., p. 306). That is why the question of word recognition after the word’s offset has long been neglected. In his opinion, semantic or pragmatic context can be responsible for a delay in reaching the word’s uniqueness point because context may influence the bottom-up process of word recognition. Grosjean’s explanation contradicts Norris (1994), who questioned whether feedback in speech recognition is necessary at all (see also Norris, McQueen, and Cutler, 2000).

Let us come back to the mentioned string *bun* which can either be a word in itself or part of another word like *bundle*. Corpus studies proved that the majority of the polysyllabic
words in English have shorter words embedded within them. Luce (1986), for instance, checked 20,000 phonetically transcribed English words in an online-database regarding their uniqueness point. He found that the majority of shorter words, which usually have a higher frequency, were not unique at their offset. Luce concluded that they “will not be recognised until some portion of the word-initial acoustic-phonetic information of the following word is processed” (Luce, 1986, p. 158). Frauenfelder (1991) obtained similar results for Dutch. Both corpus studies support the assumption put forward by Grosjean (1985).

Further evidence for context-dependent word perception comes from Gaskell and Marslen-Wilson (1996). They tested in cross-modal priming experiments if lexical access was possible when instead of the real word like wicked a phonetic variation is presented, e.g. [ˈwɪkɪb]. The perception of isolated targets was compared to the perception of targets embedded in a lexical context. Generally, targets with context were easier to recognise than isolated ones. Surprisingly, isolated perception of unchanged and distorted targets ([ˈwikɪd] - [ˈwikɪb]) showed similar priming effects. The authors found that distorted targets were perceptually acceptable when they might be embedded in a context that licensed a regular assimilation process. In our example, place assimilation is possible, for instance in the context [ˌwikɪb 'præŋk]. Gaskell and Marslen-Wilson offered two explanations:

1. The phonologically changed phonemes were treated as noise which was insufficient to disrupt lexical access.

2. The last phonemes of the targets were treated as the first segments of the following word, while the last segment of the target was perceived as deleted.

The perception of isolated, distorted targets was also examined by Connine et al. (1997). They created non-words (e.g. zeramic) that sounded nearly like meaningful words (ceramic) and found that non-words which differed in one or two phonological features were tolerated as meaningful by listeners. The effect disappeared when more features were manipulated. The presented example also touches the issue of normalisation. Basically, the speech signal of the same phoneme varies between contexts, speakers, and dialects. Each person utters the same segment differently. Even for the same speaker, the segment’s signal differs with reference to speaking rate, emotional state, microphone distance, etc. (Goldstein, 1997; Lahiri and Reetz, 2002). However, listeners are able to understand speech from various talkers, even when they are unfamiliar to them. This is possible through the application of normalisation techniques (McQueen and Cutler,
The auditory system adopts differences in the manner of speaking (Mullennix, Pisoni, and Martin, 1989). Frauenfelder and Tyler (1987) were cited in this section, who stated that the activation of lexical entries is dependent on the speech stream’s “goodness of fit”. However, most of the non-standard pronunciations forms of casual speech will nevertheless be recognised by listeners without much difficulty (McQueen and Cutler, 2010). Listeners are able to apply normalisation techniques in order to restore distorted or missing speech sounds in a sentence on the basis of the prior and following context. Usually, they do not even realise that a sound is distorted or missing. Warren and Obusek (1971) called this phenomenon “phonemic restoration” (PhR). They presented listeners with speech material in which phonemes were replaced by a non-speech sound (click, buzz, cough, etc.). The task of the listeners was to detect distorted phonemes. Warren and Obusek (1971) found evidence for a PhR-effect, because listeners stated that they had heard the target phonemes (despite the fact that the phonemes were missing from the speech stream). The authors explained that “the speech sounds synthesised through phonemic restorations [...] cannot be distinguished by the listener from those physically present” (ibid., p. 360). Even when the test persons had been informed that one or more sounds of the following sequence had been replaced, they were not able to detect which sound was missing. In the course of their experiment, Warren and Obusek used a silent interval instead of noise as the replacement for a phoneme and played the so manipulated sequences to their listeners. Other than before, listeners were quick to localise the missing phoneme and no PhR-effect was found. The authors’ explanation for this is as follows: auditory perception in everyday life is disturbed by a wide range of extraneous sounds. Therefore, the perceptive system is accustomed to deal with speech distortions like clicks or coughs. Missing information is compensated. Silent intervals, in contrast, have an important function in speech segmentation. Listeners’ ability to locate them correctly therefore reflects their ability to integrate suprasegmental information in the segmentation process (Warren and Obusek, 1971).

Warren and Sherman (1974) replicated the experiment, but changed the kind of speech sound distortions. During the recordings of the sentence stimuli, the target phonemes that were later cut out, were deliberately mispronounced, e.g. George waited for the deli[b/v]ery of his new color TV (in brackets: the phoneme that was uttered by the

---

4Warren and Obusek regard silent intervals as “suprasegmental phonemes” due to the crucial role silent intervals play in speech production and perception (Warren and Obusek, 1971, p. 361).
1. Introduction

Warren and Sherman’s aim was to erase acoustic cues that might enable listeners to restore a missing phoneme on the basis of neighbouring speech sounds. The deleted phonemes were then replaced by noise or a silent interval, and the material was played to listeners. The authors obtained the same results as Warren and Obusek (1971). Distorted phonemes were “heard”; silent intervals were detected as such.

Cole (1973) tested how far restoration techniques reached. Mispronunciations were created by changing one phoneme in a word by one, two or four distinctive features. Targets were embedded in a speech stream. Test persons were asked to detect mispronounced words. Items which had only one manipulated distinctive feature were less often detected (fewer than 30 per cent) than those changed by two or four distinctive features (60 per cent/75 per cent correct hits respectively). Cole came to the same conclusion as Warren and Obusek (1971), namely that listeners do not pay attention to all of the acoustic information that is present in the speech signal. A certain amount of “noise” is tolerated in the recognition process. Since the mispronounced phonemes had been embedded in a contextually coherent environment, a restoration process was possible.

The tolerance of distorted input is also addressed in the concept of Underspecification (Archangeli, 1988; Lahiri and Marslen-Wilson, 1991; Gaskell and Marslen-Wilson, 1996; Lahiri and Reetz, 2002). A good example for a word recognition model that is based on underspecification is the Lahiri and Reetz Featurally Underspecified Lexicon (FUL) model (Reetz, 1998; Lahiri, 1999; Reetz, 1999; Lahiri and Reetz, 2002). Assuming that each segment is lexically represented by various features to separate it from any other segment in the phonology of a particular language, FUL claims that listeners do not need to process all features to identify a segment. Redundant information is not stored in German, for instance the feature [voiced] for sonorants, since all sonorants are voiced. Hence, the feature [voiced] is redundant for this class of sounds. FUL is based on four assumptions:

1. Phonological phoneme variants are not stored; variants are abstract and underspecified.
2. Each morpheme has a unique representation; no phonological variants are listed.
3. The perception system analyses the speech stream for rough acoustic features, which are transformed into phonological features.

The feature [labial], however, remained present.
4. Phonological features are mapped on to the lexical representations. This is achieved by a three-way matching system: match, no mismatch and mismatch.

Beside the phonological information, morphological, syntactic and semantic information becomes available (Lahiri and Reetz, 2002, p. 638). A “match” happens when the signal and the lexicon have the same features. All features fit exactly, in number and quality. A “mismatch” occurs when the signal and the lexicon have contradicting features. “No mismatch” means that the segment does not fit optimally, but the candidate is not rejected (Lahiri and Reetz, 2002, p. 640).

For instance, FUL predicts that if [labial] is extracted from the signal, it matches with underlying /m/ and does not mismatch with /n/ which is underspecified. The process should not work vice versa because a [coronal] sound from the signal mismatches with stored /m/. Lahiri and Coillie (1999) showed the functionality of FUL in a priming experiment. Listeners heard words like Bahn/railway or Lärhm/noise in isolation. At the offset of the acoustic stimulus, test persons saw a semantically related target (Zug/train or Krach/bang) in the written form and had to decide whether it was a word or not. When the acoustic stimulus was a real word, the semantically related target was primed. The aim of the experiment was to test if related, acoustic non-word variants like *Bahm or *Lärhm also produced a priming effect. In accordance with the principle of underspecification, *Bahm had a priming effect while *Lärhm had not. *Bahm does not mismatch with the lexical representation of Bahn and therefore successfully activated Zug. *Lärhm, however, produced a mismatch with Lärhm and its semantic associate was not activated.

A different approach for a word recognition model was pursued by Hintzman (1986) and Goldinger (1998). Both linguistics provided evidence for the exemplar model, which is based on the assumption that “each experience produces a separate memory trace” (Hintzman, 1986, p. 411). Jäger (2008) explained that after a cue acoustic-phonetic contact, all traces are activated according to their similarity to the cue. Observation causes stored exemplars to resonate. Superimposing resonating exemplars produce echoes which are stored in the mental lexicon. The observation itself is not stored, only its echo. Variable stimuli are mapped onto a canonical representation that approximate a prototype. Hintzman and Goldinger assumed that a prototype is an area of high exemplar density. Each exemplar functions as a magnet and pulls the echo of an observation towards itself. Goldinger (1998) proved this assumption in a shadowing ex-
periment. He asked participants of a perception-production study to repeat (to shadow) words they heard via loudspeaker. Both high and low-frequency words were produced in this way. Later, other participants were asked to classify each word produced by a shadower in comparison to the original word the shadower had heard via loudspeaker. The task was to judge whether the produced word sounded rather like an imitation of the originally perceived word or not. Low-frequency words lead to stronger imitation than high-frequency words. Goldinger assumed that low-frequency words have fewer echoes and will therefore exhibit stronger effects of imitation.

Exemplar models are useful to demonstrate that frequency of occurrence plays an important role in the word recognition process. Generally, linguists agree that high-frequency words have an advantage in word recognition tasks. However, lexical frequency is only one factor among many in this complex process. Which factor becomes important at what time is not yet fully clear.

We mentioned earlier in this chapter that Warren and Marslen-Wilson (1987) asked listeners to identify words before their acoustic offset (e.g. scoop and scoot), which the test persons did successfully. As a side result, they found that correct identification worked independently from the lexical frequency of the words.

Connine, Titone, and Wang (1993) analysed frequency-related word perception in greater detail. They asked listeners to identify words containing ambiguous phonemes (e.g. best and pest where the /b/p/ phoneme was an in-between form). Here, one word was frequently used in everyday speech (best), the other one rather rarely (pest). Words were presented to listeners as acoustic stimuli, the task was to indicate the initial sound of the stimulus. Test persons preferred the high-frequency word. Connine, Titone, and Wang (1993) called this the “intrinsic frequency effect”. In a next step, they manipulated the composition of the lists in which the targets were presented. When the target words were presented in a list with high-frequency unambiguous fillers, the listeners’ bias towards high-frequency words increased (35 per cent bias compared to 13 per cent in the unmanipulated condition which was a list with mixed high and low-frequency unambiguous fillers). On the other hand, when the target words were presented in a list with low-frequency unambiguous fillers, participants gave more low-frequency responses. Connine et al. also measured reaction times and found that reaction times were affected by the frequency list type (high/low/mixed) rather than the frequency of the target words. The authors therefore concluded that “word frequency functioned as a default source of biasing information used late in the decision process rather than an early source
of information used to shape lexical hypothesis” (ibid., p. 91).

Cleland et al. (2006, p. 105) noted that the experiments of Connine, Titone, and Wang (1993) “were based on identification rather than lexical decision”. Their own aim was to find the locus of the frequency effect. For this purpose, they chose 120 low- and 120 high-frequency words from the CELEX database (Baayen, Piepenbrock, and Rijn, 1995). In addition to the targets, 240 corresponding non-words were generated by altering one phoneme in each high and low-frequency target word. All words were monosyllabic and were presented as acoustic stimuli. Test persons had to decide if they heard a word or a non-word. The decision was preceded by yet another task. Shortly before the acoustic stimulus, a coloured square (blue or green) appeared on a computer screen and had to be identified as “blue” or “green”. The authors introduced this extra task in order to address the consequences of any processing bottleneck in spoken word recognition. Test persons had positioned their hands on a keyboard. The task order was:

1. Press a button for green or blue (left hand, middle/index finger).

2. Press button for word or non-word (right hand/middle index finger).

Test persons were told to respond accurately to both tasks, but to respond quickly to the colour. Reaction patterns suggested that lexical decisions (task 2) influenced the speed of the colour decisions (task 1) depending on lexical frequency and phonetic match of the input stimulus. High-frequency words were recognised faster than low-frequency words. Cleland et al. (2006) thus found evidence of an early locus for frequency effects in spoken word recognition. The brain’s processing system can only deal with one task at a time. This reaction pattern is known as the “bottleneck constraint”.

Van de Ven, Tucker, and Ernestus (2009) also examined frequency-dependent word recognition, in particular the perception of reduced word variants. They found that common words are better recognised than uncommon words, even when the common words were presented in highly reduced variants. A native speaker of Canadian English was recorded reading a list of high and low-frequency words (1) clearly and in a normal speech tempo and (2) as quickly as possible in order to elicit reduced speech. Participants of a self-paced perception experiment listened to all stimuli and had to make lexical decisions about the target words. High-frequency words were better recognised than low-frequency words. The result was achieved for both the clearly articulated version and the reduced form. The given examples demonstrate that word frequency is an important factor in the word recognition process.
Apart from frequency-driven word recognition, categorical word perception has been assumed to play an important role. The principle proposes that speech sounds are cast into categories according to the respective language (Liberman et al., 1967). Ganong (1980) demonstrated that the category boundary shifts towards a lexical item when listeners hear a speech continuum between a word and a non-word. He recorded the non-word *tash* and manipulated its initial stop /t/ in such a way that it gradually changed to /d/. Eventually, after several in-between-forms, the meaningful *dash* emerged. Ganong played this dash-tash continuum to test persons and asked them if they heard *dash* or *tash*. As a result, listeners showed a preference for the real word, which means that they forgave phonetic vagueness of the stop as long as the resulting output was meaningful. Goldstein (1997) explained that the auditory system of humans applies a simplification method. Instead of processing a great variety of speech information that differs only slightly, the brain perceives in categories. Goldstein called the point where perception changed “phonetic boundary” (ibid., p. 402).

### 1.3. Outline of the thesis

This chapter gave background information about word production and perception strategies. Chapter 2 will provide more details about the chosen wb markers. The experimental part starts with chapter 3, the production experiment. Here, the creation of our wb related speech corpus is explained. The analysis of the speech corpus provided first results about place, frequency, duration and phonetic properties of wb markers. Chapter 4 continues the experimental part and follows the question about which wb markers are important for the recognition process. Answers were found in a two-part perception experiment. Chapter 5 summarises the findings of this thesis, evaluates the experimental design of the perception experiment, and provides suggestions for future research.
2. Theoretical background: Word boundary markers in German speech

This chapter gives the theoretical background on the investigated wb markers. These are discussed with respect to production and perception in German and other languages. Each section refers to one wb marker and starts with its definition. After that follows the discussion of literature with regard to the production and perception of the respective wb markers relevant for this thesis. Despite the fact that our production experiment considers only unpredictable wb markers, which excludes /l/-allophones and stress (see chapter 1), this overview will also provide some background on the production of these two markers. Research results obtained for German speech will usually be introduced first, since it is the examined language in this thesis. Works on other languages will follow. Within the frame of our experiments we have analysed the influence of dialect on wb perception. The last section of this chapter will therefore provide some theoretical background on dialect specifics.

2.1. Glottal stops/creaky voice

A glottal stop is produced when the vocal folds are tightly closed and suddenly released. Often, the glottal stop is paired with another laryngeal gesture, creaky voice. If a vowel or sometimes a sonorant consonant is uttered with vocal folds too tight to produce a real vibration when the air stream passes through them (or too slack in the case of speakers with a low-frequency bass line), the sound comes out creaky, hence the term creaky voice. Only a small part of the vocal folds vibrates. The spectral analysis shows irregular glottal pulses and an interrupted or no fundamental frequency line at all.

In this thesis, glottal stops are of central interest and creaky parts were only regarded
as their supplements. Word-initial vowels are a frequent location for the glottal stop\(^1\), a fact that makes the feature interesting as a \(\text{wb}\) signal. In German speech, however, it also appears in another position frequently, namely word-medially substituting a stop while schwa is deleted (Trubetzkoy, 1939; Kohler, 1995), for example \(\text{könnten/could} \ [\text{k}^b\text{öntn}] \rightarrow [\text{k}^b\text{öen}?n]\).

This leads us to the question whether the glottal stop is part of the phonemic inventory of German or not. Wiese (2000) affirmed this and classified the glottal stop as a consonant for which no particular oral-articulatory gesture is defined. In Wiese’s opinion, the glottal stop can be described by the feature “constricted glottis” within the feature matrix which will be exclusively allocated to \([\text{?}]\). The special position of the glottal stop as a phoneme becomes clear when the structure of German syllables is considered. Not every spoken syllable has to contain a vowel (Wiese, 2000). This principle has been demonstrated with the previously given example word \(\text{könnten/could}\) which has the canonical form \([\text{k}^b\text{öntn}]\), but might as well be pronounced \([\text{k}^b\text{öen}?n]\). In Wiese’s theory, \([\text{?}]\) fills an empty onset node\(^2\). Kohler (1995), on the other hand, argued that the glottal stop is no phoneme, but part of a vowel onset or a stem morpheme onset. Its production is not mandatory and its absence does not lead to ill-formed words. Kohler also described the word-medial glottal stop as a reduction phenomenon of supraglottal stop articulation (Kohler, 1994). Köser (2001) supported this assumption with her production study of alveolar stops /t/ and /d/ in word-medial position. She examined words which had the structure nasal/lateral + plosive + /\(\text{a}\)/ + nasal/lateral. Glottalisations were common substitutions for the two examined plosives and /\(\text{a}\)/ was deleted in the substitution process.

Glottal stops are often articulated in conjunction with creaky voice. The constriction and release of the vocal folds provide a productive condition for a creaky part to follow. As an isolated feature, however, creaky voice seems to have a weaker \(\text{wb}\) signalling capacity when compared to glottal stops. There are two reasons. Firstly, all sonorants have the general potential to be produced creaky, independent of their position in a word (Gordon and Ladefoged, 2001). Secondly, speakers with a low-frequency bass often show a distinct creaky voice articulation, as illustrated in Figure 2.1 even though it can occur with any pitch. When excluding the speaker’s individual voice characteristics, creaky voice has yet preferred locations. At the phrase level, both phrase-initial and -final segments are prone to creaky voice articulation (Di Napoli, 2011).

\(^1\)Occasionally, word-initial consonants are also preceded by a glottal stop (Kohler, 1999).

\(^2\)Wiese assumes that German has no onset-less syllables.
Infante Rios and Perez Sanz (2011) observed that there is a difference between a creak and creaky voice. They argued that creaks (pulse registers) usually appear at the end of an utterance, while creaky voice can be found at various positions within the utterance. Infante Rios and Perez Sanz proved their point through results obtained by electroglottography (EGG, a non-invasive method, which provides information on vocal fold vibration through electrodes placed on the neck at the approximate position of the vocal folds) in addition to perceptual and spectrographical analyses. Six male speakers were recorded producing spontaneous speech. Creaks and creaky voice segments were found to be very similar sounds in the signal, distinguishable only through the EGG-waveform patterns. Creaks (pulses), which had a typical utterance-final position, showed non-periodic cycles. Creaky voice produced semi-periodic cycles and appeared within the utterance. In theory, creaks might function as a wb marker, because they signal the end of an utterance. In practice, the measurement method as proposed by Infante Rios and Perez Sanz (2011) is rather complex and could not be applied during the production experiment for this thesis.

After this brief description of the characteristics of glottal stops and creaky voice, we take a closer look into their application during the process of speech production. The literature presented in the following refers to the question, how reliably glottal stops and creaky voice segments occur at the beginnings or ends of words. They are discussed as individual and combined wb markers.
2. Theoretical background: Word boundary markers in German speech

2.1.1. Production

The literature we refer to in this section, does not always differentiate clearly among the possible glottal gestures and summarises glottal activity under the terms “glottalisation” or “laryngealisation”. Consequently, results may be interesting for this thesis in general, but not always comparable to studies that explicitly refer to glottal stops and creaky voice.

Krech (1968) examined 10,000 German words with initial vowels and found that 41 per cent of her targets were realised with a glottal stop, 27 per cent without any laryngeal gesture at all (clitisations, assimilations etc.) and 14 per cent with a soft onset where the vocal folds softly start to vibrate (“weicher Einsatz”) without a preceding closure. Other forms were breathy or pressed onsets, creaky voice onsets or mixed forms, which were found in less than 10 per cent of the examined speech material. Combinations of glottal stops and creaky voice were not explicitly looked into. Krech also followed-up the issue of speech tempo. She tested whether an increase in the tempo of speech influenced the frequency of fully realised glottal stops or not. With a higher speech tempo, the quality of the uttered sequence usually changes. Vowels are shortened, assimilation processes become more frequent, phonemes are dropped. Krech recorded 72 professional German speakers who spoke either slowly (programme announcers) or fast (news readers). Both groups of speakers articulated glottal stops with equally good quality. Fast and slow speaking test persons produced nearly the same number of fully realised glottal stops. Krech explained that glottal stops are border signals and have therefore to be pronounced properly in order to maintain the comprehensibility of the text. Her observation reflects the important position of glottal stops within German phonology. Krech (1968) also analysed her speech corpus regarding the context preceding glottal stops. She found that after speech pauses, which often coincide with intonational phrase (IP) boundaries, there is a strong tendency to articulate vowels with a glottal stop - no matter if the vowel belonged to a stressed or to an unstressed syllable. Kohler (1994) investigated glottal stops and glottalisations at word boundaries in connected speech and obtained similar results. He offered two explanations for the increased occurrence of glottalisations at IP boundaries. The first one is that glottalised phonemes appear naturally when producing a vowel onset after a pause due to the speech organs’ mechanical constraints. The

---

Krech (1968) examined 10,000 German words with initial vowels and found that 41 per cent of her targets were realised with a glottal stop, 27 per cent without any laryngeal gesture at all (clitisations, assimilations etc.) and 14 per cent with a soft onset where the vocal folds softly start to vibrate (“weicher Einsatz”) without a preceding closure. Other forms were breathy or pressed onsets, creaky voice onsets or mixed forms, which were found in less than 10 per cent of the examined speech material. Combinations of glottal stops and creaky voice were not explicitly looked into. Krech also followed-up the issue of speech tempo. She tested whether an increase in the tempo of speech influenced the frequency of fully realised glottal stops or not. With a higher speech tempo, the quality of the uttered sequence usually changes. Vowels are shortened, assimilation processes become more frequent, phonemes are dropped. Krech recorded 72 professional German speakers who spoke either slowly (programme announcers) or fast (news readers). Both groups of speakers articulated glottal stops with equally good quality. Fast and slow speaking test persons produced nearly the same number of fully realised glottal stops. Krech explained that glottal stops are border signals and have therefore to be pronounced properly in order to maintain the comprehensibility of the text. Her observation reflects the important position of glottal stops within German phonology. Krech (1968) also analysed her speech corpus regarding the context preceding glottal stops. She found that after speech pauses, which often coincide with intonational phrase (IP) boundaries, there is a strong tendency to articulate vowels with a glottal stop - no matter if the vowel belonged to a stressed or to an unstressed syllable. Kohler (1994) investigated glottal stops and glottalisations at word boundaries in connected speech and obtained similar results. He offered two explanations for the increased occurrence of glottalisations at IP boundaries. The first one is that glottalised phonemes appear naturally when producing a vowel onset after a pause due to the speech organs’ mechanical constraints. The

---

Before TV magazines were published, the TV station’s programme for the following week was read every Sunday morning by a so-called programme announcer. This announcer spoke very slowly in order to enable the TV audience to write the programme down.
second one assumes that a pause – which is a strong boundary signal in connected speech – attracts other boundary signals such as a glottalised vowel onset. They are therefore associated. Trouvain (2004) agreed with the latter explanation; he described glottal stops as one indicator for prosodic breaks in German speech.

Another study of glottal stops and creaky voice segments in German was presented by Wesener (1999). He analysed German vowel-initial indefinite articles (ein, eine, etc.) using the Kiel Corpus of Spontaneous Speech (as published on CD-ROM, IPDS, 1995, 1996, 1997) and the Kiel Corpus of Read Speech (as published on CD-ROM, IPDS, 1994) and found that, in general, vowel onsets were rather marked by creaky voice than a glottal stop. This result differs from Krech’s corpus study in which creaky voice in initial vowels appeared rather rarely. Wesener also reported that if a glottal stop occurred, then mostly in connection with creaky voice. A glottal stop without creaky voice was in fact the least common realisation. The differing results of Krech (1968) and Wesener (1999) regarding the frequency of creaky voice at vowel onsets can have various reasons: Krech analysed more speakers than Wesener, all of her participants were professionals while the Kiel Corpus – Wesener’s source – consists of speech contributed by untrained speakers. Wesener included spontaneous speech data, Krech only looked at read speech. Wesener’s results are based on limited material (indefinite articles), Krech investigated a greater variety of words.

The functionality of glottal stops and creaky voice as wb markers also holds for other languages. The findings for German are mirrored in studies for English. Dilley, Shattuck-Hufnagel, and Ostendorf (1996) examined 3709 word-initial vowels and found that all speakers produced significantly more glottalisations at the beginning of a new IP. They observed that even reduced vowels in word-initial position were likely to be produced as glottalised when they started an IP. Full vowels, however, have been found to glottalise more often in this place. Pierrehumbert and Talkin (1992) had reported similar results earlier. An acoustic-phonetic study of Finnish by Lehiste (1965) substantiated the wb signalling capacity of glottal stops. She recorded word pairs like lintuansa/bird, partitive and possessive suffix, 3rd p. and lintu-ansa/bird-trap in order to compare the boundary

---

4There are some special cases in German like Theatre/theatre → [ˈθiːtə:ɹ] (example from Wiese, 2000, p. 59) or Michael/first name → [ˈmɪcaːl] (example from Kohler, 1995, p. 168). If a word-medial syllable begins with a vowel and receives stronger stress than the preceding one, it might be realised with a glottal stop (Wiese, 1995). The number of these words is negligibly small. Prefixed nouns and verbs, however, represent a large group of words with glottal stops at morpheme boundaries (beiden/to swear to something → [bəˈnaːdən], Erörterung/conquest → [ɛrˈɔːtɐrʊŋ], the respective prefixes are underlined). They were not included in this thesis.
section of the compound to the corresponding area of the declensed noun. Speakers often inserted a glottal stop or a short period of laryngealisation between two vowels of the compound *lintu-ansa* while the declined noun *lintuansa* was produced without a wb marker. Lehiste (1965) also mentioned the application of other wb markers such as final segment lengthening in the first word of the compound. Hence, in order to disambiguate phonemically identical word pairs, speakers might have produced more than one wb marker to make the boundary stronger.

2.1.2. Perception

Word-initial glottal stops and creaky voice segments are regarded as well-perceptible wb markers. Exemplarily, the perception experiment conducted by Nakatani and Dukes (1977) is presented here to illustrate this. The authors tested which prosodic cues influenced the perception of word junctures and recorded phrase pairs like *no notion* and *known ocean*. These “parent phrases” were cut into four portions (*no*, *notion*, *known*, *ocean*) and cross-spliced. Listeners heard the thus formed hybrid phrases and had to decide whether they heard *no ocean*, *no notion*, *known ocean* or *known notion*. When the test persons perceived a glottal stop followed by a laryngealised /o/, the overwhelming majority decided for a phrase with the second word being *ocean*. Nakatani and Dukes concluded, that “[a] glottal stop and laryngealisation (creaky voice) at the onset of a word-initial vowel were the most prominent acoustic cues at the beginning of a word” (Nakatani and Dukes, 1977, p. 719).

The presented studies in this section agree about the good wb marking potential of glottal stops and creaky voice. In the following section, we turn to stop allophones and summarise how their usability as wb markers in German and other languages has been evaluated by researchers so far.

2.2. Stop allophones

Phonetic variations of a phoneme are called allophones. They are not associated with a difference in meaning. However, if certain allophones appear only word-initially or -finally, they may be able to signal a wb. The following section summarises the wb
signalling capacity of stop allophones which is rooted in the circumstance that they can have different degrees of aspiration depending on their position in the word and on the phonemic environment.

### 2.2.1. Production

Kohler (1995) gave a comprehensive overview of the behaviour of German [+ fortis] stops /p, t, k/, as well as [- fortis] stops /b, d, g/, and their allophonic variations. Word-initial [+ fortis] stops appear with aspiration which is stronger the closer the stop stands to the stressed vowel. Kohler specified that aspiration is stronger before the stressed vowel than afterwards. Stops may also be aspirated in word-final position, but to a lesser degree compared to the initial position. In his example tut/does → [tʰu:tʰ], both stops are aspirated, but the word-initial one before the stressed vowel more so than the word-final one.

Due to final devoicing in German, word-final /b, d, g/ are uttered as /p, t, k/ and behave like [+ fortis] stops, e.g. Rad/wheel → [ra:tʰ]. German [- fortis] stops /b, d, g/ are only truly voiced in an intervocalic context, e.g. mager/skinny. At the onset, they are more or less devoiced. They differ from /p, t, k/ by means of aspiration. The initial stop in Bier/beer is unaspirated while Pier/pier is aspirated (Kohler, 1995, p. 81 ff.).

German [+ fortis] stops show a special behaviour in certain phonemic environments (ibid.). We will utilise this fact in our perception experiment.

1. Stops carrying the feature [+ fortis] are hardly aspirated in fricative-stop clusters in onsets, e.g. /ʃt/ in Stiere/bulls, and the closure duration is longer.

2. If /l/ follows after the [+ fortis] stop, as in Klippe/cliff, the sonorant is voiceless and the stop’s aspiration is either reduced or eliminated.

3. If the [+ fortis] stop is followed by nasal /n/, e.g. in Knoten/knot, the closure of the stop is often (nasally) released with little or no aspiration and the nasal is devoiced.

How do stop allophones behave in other languages? Let us have a a look at English. Church (1987) argued that the two phrases gray twine and great wine contrast with regard to their phonetic realisation, because word-initial /t/ in twine is aspirated while word-final /t/ in great is not. Word-initially, the explosion that occurs when the closure
of a stop is released, shows higher intensity values than the word-final allophone. Pressure in the initial position is greater and the release is fully executed, while word-final stops are produced with lower initial pressure and the release is hardly or not at all perceptible. This is why the /t/-allophone signals which sequence is uttered: *gray twine* or *great wine*.

The degree of aspiration that accompanies the [⁺ fortis] stop is not only dependent on word position, but also on its place of articulation (Peterson and Lehiste, 1960). They measured all English [⁺ fortis] stops and found that aspiration after /p/ was shortest (58 ms), followed by /t/ (69 ms) and /k/ (75 ms) showing the longest period.

2.2.2. Perception

In the previous section, we cited Kohler (1995) who explained that German stops can generally be aspirated in all word positions, but aspiration is stronger before the stressed vowel. If there is a position-induced difference in strength, position-interchanged stop allophones (e.g. a highly aspirated one in the word-final position) might be perceptible as a “wrong” allophone. In our perception experiment, we place the “wrong” stop allophones in initial, medial and final word positions. For instance, we added a fricative as the initial segment to a word starting with a [⁺ fortis] stop, e.g. /Stiere/animals. The resulting sequence [Stirə] is a meaningful word, but the acoustic target carries a highly aspirated stop allophone in a position that is little or not at all aspirated in the natural word *Stiere/bulls* [Stirə]. The question is if listeners are able to hear the wb signal, namely the “wrong” stop allophone. More details will follow in the perception section 2.2.2.

In the case of fricative-stop clusters, there is another factor apart from aspiration that determines if a real word or a phonemically identical sequence is perceived, namely the duration of the fricative preceding the stop. Shatzman and McQueen (2006) showed this in an eye-tracking experiment with Dutch participants. The authors used acoustic stimuli comprising of stops preceded by cross-spliced /s/, e.g. target: *eens pijp (once pipe)*, competitor: *spijker (nail)*, which were in part ambiguous with respect to the first syllable (*spij*). Note that initial stops are unaspirated in Dutch, aspiration did therefore not present a cue for discrimination (van Alphen and Smith, 2004), frictive duration, however, did. Target and competitor were embedded in sentences: *Ik zou ooit eens pijp willen roken/I would like to smoke a pipe once* and *Ik zou ooit een spijker willen kopen/I would like to buy a nail*. While listening to the sentences, the test persons saw
2. Theoretical background: Word boundary markers in German speech

four pictures appearing on a screen in front of them showing the target (a pipe), the competitor (a nail) and two unrelated items (e.g. a diving mask and a feather). The participants’ eye-movements were monitored. Frication duration of /s/ was either long (103 ms average duration) or short (73 ms average duration). When /s/ was long, the word with the initial consonant cluster (spijker) was favoured. The conclusion we draw from these results is that the influence of one wb marker (here: stop allophones in word-initial fricative-stop clusters) on word segmentation can only be tested in a perception experiment if other wb markers (here: fricative duration) are manipulated in such a way that they cease to give additional cues to the listeners. Otherwise, experimental results might be unreliable.

Another study that provides interesting insights into how stop allophones are perceived, is the one by Lotz et al. (1960). They tested the perception of American English stop consonants /p/, /t/, /k/ which had originally been part of an /s/-stop-cluster (spill, still, skill). After deleting the /s/-segment, the remaining sequences were played to listeners. Like in German, [+ fortis] stops are usually pronounced with aspiration in word-initial position, while in fricative-stop-clusters, aspiration of the stop disappears. Unsurprisingly, American listeners reported to hear words beginning with voiced stops (bill, dill, gill) all of which are meaningful words. The authors concluded that “the lack of aspiration is a dominant cue for forcing the evaluation of the stop in the direction /b, d, g/” (ibid., p. 77). These findings further support our assumption that the presence vs. absence of aspiration in stop allophones may signal a wb.

While the signalling potential of stop allophones is founded in aspirational cues, /l/-allophones draw this capacity upon their changing spectral properties in different phonetic environments. The next section gives the background for this proposition.

2.3. L-allophones

The usability of /l/-allophones as wb markers arises from the fact that their formant transitions and frequency spectra depend on the surrounding phonemes. A palatalised /l/, for instance, that is exchanged by the /l/ stemming from the context die Pappe liegt/the cardboard lies might render the sound of the word unnatural.

5Possible variants with [+ fortis] stops had also been meaningful: pill, till, kill.
2. Theoretical background: Word boundary markers in German speech

2.3.1. Production

Generally, the German lateral /l/ is a voiced coronal (Wiese, 2000). The sound is formed when the blade or the tip of the tongue touches the alveolar ridge, while the sides have no contact with the palate so that the airstream can escape through the two gaps (Hall, 2003, p. 56). Since there is no clear phonological rule for /l/-allophone generation in German, the variants of /l/ rather reflect fine phonetic detail. However, for the sake of readability, we will continue to use the term allophone.

Allophonic variations of /l/ in German are brought on by the neighbourhood of the liquid. During the articulation of the lateral, tongue position, as well as the front-back movement of the tongue’s body and the opening angle of the mouth adjust to the surrounding phonemes. It is also possible that the lateral influences the place of articulation of the preceding phoneme. This is for instance the case when /l/ follows a stop (e.g. Klippe/ cliff). What we see is a coarticulation process, which usually leads to a palatalised stop which is no longer aspirated (Kohler, 1995).

While German has no clear allophonic rules for /l/, other languages have. In English, for instance, the phoneme /l/ appears in two allophonic variants - as a word-initial light form, e.g. lime, or as a dark variant in word-final position, e.g. call (Church, 1987). They do not contrast, because they cannot appear in the same context: light /l/ never appears word-finally, and the dark form never appears word-initially (Reetz and Jongman, 2009; Hall, 2000).

Apart from the allophonic rule for /l/, the English lateral may also be erased or vocalised. Some Commonwealth English variants display position-induced /l/-deletion. Speakers of Australian English, for instance, do not utter the lateral if it follows a low or back vowel. The pre-lateral vowel is lengthened (Proctor, 2011). Examples are:

\[
calf \rightarrow [k\v'f]
\]

\[
almond \rightarrow [\v:mand]
\]

(Examples taken from: Proctor, 2011)

Laterals in post-vocalic position are also prone to assimilation processes in some dialects of Northern England. Here, laterals in the mentioned location will be vocalised to a mid-back vowel: milk \(\rightarrow [m\v'k], [m\v k]\) (Hardcastle and Barry, 1989).

---

6 The phenomenon of coarticulation is resumed in section 3.2.3
2. Theoretical background: Word boundary markers in German speech

2.3.2. Perception

Let us have a closer look at the example mentioned in the introduction to this section. The word Pappel/poplar and the phonemically identical sequence taken out of the utterance die Pappe liegt/the cardboard lies should be discriminable by listeners, because the final /l/ sounds natural in the first and unnatural in the second example. The reason, why the final /l/ in Pappel[liegt] sounds “wrong” is that tongue and lip positions have already adjusted to the following vowel /i/.

The perception of targets with /l/ is sometimes difficult, especially if the lateral is in word-final position. There is a likelihood of confusion with rhotics. Frick (2003) demonstrated that in a gating experiment with German listeners. Frick’s participants heard sentences which unfolded phoneme by phoneme. Consequently, the acoustic stimulus expanded gradually. After each new phoneme that was added to the sound chain, test persons should write down the word/words/sentence they thought is developing. Some of the target words contained the lateral at a word-final position. The lateral was sometimes perceived as a rhotic if both resulting strings were meaningful and context was yet insufficient (e.g. vier/four instead of viel/much). In sentence-initial position, /l/ was well recognised. This might be attributed to the phenomenon of phrase-initial strengthening (Nakatani, 1978). In this process the phrase-initial segment is lengthened. The longer duration of /l/ might have lead to a better distinguishability of the lateral.

The misperception of laterals as rhotics and vice versa is also known from other languages with two phonemic liquids such as Greek (Müller, 2010). Interestingly, the confusion disappears when test persons have to listen for /l/ and /r/ in nonsense sequences. Koo and Cole (2006) experimented with liquids and rhotics in non-meaningful strings and demonstrated that adults were able to learn phonotactic dependencies of a fictional language. Among the phonotactic dependencies were liquid assimilation and dissimilation. Test persons first underwent a study phase in which they heard and repeated nonsense words that initiated the phonotactic constraint of liquid assimilation, e.g. salile - (C[onsonant]2 and C3 are identical liquids) or liquid dissimilation, e.g. salire (C2 and C3 are different liquids). After the training phase, they played targets to the listeners and asked them to distinguish legal from illegal words which the test persons mastered.

While two two studies are not directly comparable, the question remains why listeners performed better with non-meaningful than with meaningful speech. The answer is

7A direct comparison also demands schwa-deletion in Pappel[liegt].
probably rooted in the complexity of a natural language. When processing meaningful sequences, the auditory input is constantly compared to the entries in the mental lexicon. Word candidates are activated according to the incoming phonetic/phonological information. The fictional language processed by the listeners of Koo and Cole (2006), however, followed some simple rules. Only a few words had to be stored in the mental lexicon. Hence, participants could concentrate on the task of word legitimacy judgement. There was not a lot more of other information to process.

In this section, we have discussed the production and perception of /l/-allophones and how they might signal a wb. The following section summarises the functionality of silent intervals as wb markers in connected speech. Some linguists use other terms like “speech pause” or “silence” which are retained in the discussion of their research.

### 2.4. Silent intervals

Silent intervals can be found at all places of an utterance. They are predictable to a certain extent. For instance, speakers tend to take breath - and thus insert a silent interval - at syntactic boundaries (Lieberman, 1967). However, the area we are interested in is the word level. The questions we follow-up are the following: Do speakers insert a silent interval for the sake of unambiguousness of phonemically identical, but syntactically differing sequences, for instance between gut and haben to distinguish the phrase [es] gut haben/being well off from the noun Guthaben/credit balance? And if they do, will listeners be able to use this information for disambiguation when the two sequences are heard in isolation?

An interesting situation arises when a wb and a stop closure fall together (e.g. in Koda_k[lingt]/coda sounds, the underscore shows the wb as opposed to Kodak/name of a company). In both sequences, there is some portion of silence after vowel /a/. The question is if the silent interval in Koda_k is longer, because the stop closure coincides with the wb.

Trouvain (2004) also mentioned “perceived” silence. Syllable lengthening, for instance, is often interpreted as a silent interval, which might be due to the fact that both features often appear in unison at the beginning or at the end of IP boundaries.
2.4.1. Production

One characteristic of speech is that silent intervals at the word level appear in unusual places (focus, hesitation), also within words (Ulbrich, 2005), and they are very difficult to predict. Butcher (1981) analysed which word types follow after a speech pause in German. He reported that more pauses are to be found before function words (especially before connectors like und/and and oder/or) than before lexical words. The reason may be that function words are often located at **IP** boundaries which are also the common places for a speech pause. Within the **IP** the number of silent intervals before lexical words and before function words was nearly equal.

Trouvain (2004) tackled the question of speech pause duration. His aim was to predict the length of speech pauses through the evaluation of prosodic markers like pitch accent or segment duration. He asked the participants of a production experiment to read a text in three self-chosen tempi: normal, fast and slow. The examined language was German. While he found general tendencies like reduction as well as a decrease of the number of pauses and prosodic breaks with a growing speech tempo, no predictable pattern for short, medium and long pauses could be derived.

Another uncertainty we have to deal with is the fact that silent intervals are part of the normal articulation process. Stop closures, for instance, might reach considerable durations. A study of German stops based on the Kiel Corpus of Spontaneous Speech (IPDS, CD-ROMS 1995, 1996, 1997) reveals maximum closure durations between 211 and 984 ms\(^8\). We therefore assume that deliberately inserted silent intervals and articulation-induced silent intervals share a large overlapping area of duration (cf. de Pijper and Sanderman, 1994).

<table>
<thead>
<tr>
<th>Stop</th>
<th>Number</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>/t/</td>
<td>11702</td>
<td>43.13</td>
<td>457.31</td>
</tr>
<tr>
<td>/d/</td>
<td>7464</td>
<td>38.31</td>
<td>983.56</td>
</tr>
<tr>
<td>/k/</td>
<td>2996</td>
<td>51.50</td>
<td>403.50</td>
</tr>
<tr>
<td>/g/</td>
<td>2435</td>
<td>45.88</td>
<td>430.19</td>
</tr>
<tr>
<td>/p/</td>
<td>1311</td>
<td>63.75</td>
<td>261.50</td>
</tr>
<tr>
<td>/b/</td>
<td>2986</td>
<td>48.09</td>
<td>211.25</td>
</tr>
</tbody>
</table>

**Table 2.1.** Analysis of stop duration (ms) by label. Source: Kiel Corpus of Spontaneous Speech (as published on CD-ROM, IPDS 1995, 1996, 1997).

\(^8\)We thank Henning Reetz for contributing the data.
Let us now turn to another language - English - and look at studies that are connected with our research topic. Gee and Grosjean (1983) analysed speech in order to examine if pauses and their relative durations were predictable. They used read speech that had been delivered at different speech tempi and developed a complex algorithm involving syntactic and prosodic features that enabled the prediction of pause placement and pause duration relative to other pauses of the utterance. One major finding was that a pause was longer the more other wb markers co-occurred at a break. Wijk (1987) re-analysed the results reported by Gee and Grosjean (1983) and concluded that prosodic features already contained sufficient information in order to predict pauses. Phonological words have to be kept as an uninterrupted unit, so that the next possible pause can be inserted between content words within a phonological phrase. Larger pauses are to be expected at phrase boundaries. Wijk mentioned that it is difficult to distinguish between phonological phrases and intonational phrases because, for the latter, linguists have not yet agreed about a definition, but see e.g. Nespor and Vogel (2007) and Ladd and Selkirk (1986) for discussions.

In a more recent study, Ramanarayanan et al. (2009) pointed out that grammatical pauses are significantly longer on average than ungrammatical ones. Grammatical pauses are considered as planned, ungrammatical pauses as unplanned breaks. The researchers used real-time magnetic resonance imaging and scanned the test persons’ oral tract during spontaneous speech production. In their experiment, Ramanarayanan et al. (2009) asked speakers to produce either read or spontaneous speech and then monitored the movement and interaction of the participants’ speech organs. Applying a specific algorithm that used criteria like jaw angle, articulator position etc., the speed of the articulators was evaluated. Ramanarayanan et al. concluded that grammatical pauses are part of the phonetic plan (cf. Levelt, 1993) while ungrammatical ones are not. Their results also show that silent intervals become more reliable as words boundary markers at certain sentence positions.

2.4.2. Perception

How do silent intervals influence speech segmentation? The research of de Pijper and Sanderman (1994) provides answers to this question. They recorded Dutch speakers who read texts. The recorded audio sequences were played to listeners who had a print-out of the texts in front of them. Listeners were asked to mark the places where they
perceived an audible boundary. They should also score the strength of each boundary using a 10-point scale. Afterwards, the authors measured the number and duration of all silent intervals in their production data and compared them with the data obtained in the perception experiment. A silent interval exceeding 100 ms resulted in high values for a prosodic boundary. The assumption that longer silent intervals indicated stronger boundaries was not found to be true. Larger durations (>200 ms, >300 ms) did not differ among each other significantly with respect to perceived strength.

While de Pijper and Sanderman (1994) showed how silent interval duration affects boundary perception, Repp et al. (1978) demonstrated that silent intervals also play an important role in speech segmentation. He carried out a perception study with American English compounds built from four words - gray, great, ship and chip - which listeners heard in all adjective-noun combinations possible (gray ship, great ship, gray chip, great chip). The insertion of a silent interval between gray and ship had no influence on the perception of the fricative in ship, but it supported the perception of word-final /t/ in gray so that listeners decided for great ship instead. That effect was reached when the silent interval was approximately 100 ms long. Apparently, the inserted silence was interpreted as the closure time of an unreleased stop.

The presented production and perception studies suggest that boundary-marking silent intervals should rather be expected at important prosodic places like IP boundaries than between words within an IP. However, the finding of de Pijper and Sanderman (1994) according to which Dutch listeners perceived a silent interval > 100 ms as a break indicator, is an interesting result which gives a good orientation for our perception experiment. However, stop closures reach similar durations, which might complicate wb identification, especially, when targets with a stop in the wb area are heard in isolation. After this summary regarding the contribution of silent intervals to speech segmentation, we come to the next wb marker - stress.

2.5. Stress

The complexity of stress is reflected by the variety of additional expressions used to describe the feature: accent, accentuation, prominence, emphasis, salience, intensity, force (Fox, 2000, p. 114). In any case, the units each term relates to are the vowels.
A \textit{wb} marking function of stress arises from the fact that the stress location directly influences word segmentation.

For example:

\begin{verbatim}
[\textit{aufbau}] \rightarrow \textit{Aufbau/construction}
[\textit{auf\,bau}] \rightarrow \textit{auf Bau [und.\,acht]/be sure to pay attention to the build-up and.\,}
\end{verbatim}

Depending on the position of stress, the string is segmented differently, and the resulting sequences are associated with different meanings. Kohler (1995) remarked for German that in every word one syllable distinguishes itself from all other syllables through the phonological feature “stress” which is characterised by a higher degree of phonetic prominence of the syllable (ibid., p. 114). Kohler (1995) also addressed the matter of larger units in connection with stress distribution. He stated that there is a tendency to stress the penultimate syllable in non-derived and non-inflected words under the condition that this syllable contains a vowel which allows stressing, e.g. \textit{H"o\,lunder/elder, \textquoteright Gro\text{\-}bian/ruffian, Franc"{i}ziska/first name} (ibid., p. 168). In a compound, word stress of both components remains while one of the words subordinates stress-wise. Former primary stress is reduced to secondary stress due to derivation in compounds (Giegerich, 1983). The result is one component with primary and one with secondary stress (Kohler, 1995, p. 188). When the first part of the compound dominates the second one, the second word might still be stressed, e.g. in \textit{\textquotesingle Rosen\,'montag/Carnival Monday, but it also works the other way round: \textquotesingle Pfingst\,'ferien/Whitsun holidays}. The difference between the two ways of stressing is ruled by semantics. In the first example, specific days of globally fixed time periods are extracted - not Carnival Tuesday, but Carnival Monday. In the second example, the category term is specified - not any holidays, but those around Whitsunday (ibid., p. 189).

Word stress is different from sentence stress (pitch accent). Note that both might interact, which is important in connection with \textit{wb} recognition. The compound \textit{\textquotesingle Feuerwehr\,'m"{a}nner/firemen}, for instance, carries primary stress on the first syllable and secondary stress on the penultimate syllable. If the same string is taken out of the sentence

\begin{verbatim}
Dass in der \textit{\textquotesingle Feuerwehr\,'M"{a}nner} arbeiten, wei\ss\ jeder/Everyone knows that men work in the fire brigade
\end{verbatim}

we will get a sequence that carries in fact two primary stresses, one for each word. While the primary stress on the first syllable in \textit{Feuerwehr/fire brigade} remains, the focus of
the sentence is on Männern/men. Hence, the focused word Männer receives the sentence stress. Syllable män carries the nuclear accent and is perceived as the most prominent one, while Feuerwehr carries the prenuclear accent on first syllable feu. The perception experiment follows-up the question, how such a sequence is segmented when heard in isolation.

Apart from word and sentence stress, the differentiation between the phonological and the phonetic level of stress is an important issue that has to be considered. Kohler (1995) has been cited before in connection with phonological parameters that define primary and secondary stress. Wiese (2000) described stress as prominence relations between phonological units as perceived by a speaker of a language (ibid., p. 272). This more general definition is adopted for our thesis.

2.5.1. Production

This section starts with the discussion of studies regarding the phonetic parameters of stress and continues with metrical/phonological issues. Most researchers who carried out production experiments in order to test which factors influence the acoustic realisation of stress, usually verified their production results with perception data. In favour of continuity, those experimental results are not split between the production and perception section but presented here as a whole.

Let us now turn to the phonetic factors that contribute to stress realisation. Lehiste (1970, p. 106) explained that stressed syllables differ from unstressed ones in terms of “greater effort” that is used to produce stressed syllables as compared to unstressed ones. She has analysed the acoustic manifestation of stress in different languages (e.g. Russian, English, Yiddish) and found that there is no one-to-one correspondence between stress and any single acoustic parameter. Stress divides the speech stream into units; it has organisational function. The smallest unit that may carry stress is the syllable, but “even if stress is associated with a monosyllabic word, consisting of a single vowel, phonetically stress is not the property of a single segment” (ibid., p. 147). Stressed and unstressed syllables can only be distinguished in comparison to a greater unit. The minimal unit for contrasting stress has to consist of two syllables (ibid., p. 147). Lehiste concluded that the acoustic realisation of stress involves the three suprasegmental parameters durational pattern of the utterance, fundamental frequency of the voice, and signal amplitude. Wiese (2000) stated that four, not three parameters contribute to the
2. Theoretical background: Word boundary markers in German speech

Effect of stress, namely pitch, duration, intensity, and quality. The parameters interact to varying degrees (ibid., p. 272). Rusch (1991) explained that there is no correlation between stress and energy distribution in German speech.

Fry (1955) claimed that duration and intensity are the physical correlates of stress. He compared phonemically identical pairs; targets were taken from sentences which had been read by test persons speaking American English. The contrasted pairs differed in their stress patterns like 'permit (noun) and per'mit (verb). Duration and intensity of the vowels were measured. Fry reported that speakers realised word stress through a longer vowel duration and a greater vowel intensity. He confirmed his results in a perception test. Vowel durations and intensity of the vowels were manipulated and various synthesised forms played to listeners. When both parameters operated in the same direction - meaning the vowel being long and of high intensity - listeners agreed that the vowel was stressed.

Hayes (1995) approached the issue of stress from a metrical point of view. His Metrical Stress Theory is based on the assumption that the type of foot gives a crucial prosodic cue of a language. In Standard German (SG), all lexical words need one stressed syllable as compared to function words which may be completely stressless. As mentioned before, stress frequently falls on the penultimate syllable. Domahs et al. (2008) explained that the preference of German for penultimate stress does not automatically justify to speak of a “default” stress pattern. The German vocabulary has a great number of bisyllabic words ending in a reduced syllable. This is the reason why penultimate stress is so frequent (ibid., p. 4). The authors rejected suggestions for a regular stress assignment to the penultimate syllable and referred to the CELEX-based corpus study of German done by Féry (1998). She found that 73 per cent of all bisyllabic German words had indeed stress on the first syllable. The picture changed when only those words with two full vowels were considered. Here, 61 per cent of the existing 1495 words had stress on the final syllable.

However, the strong-weak syllable construction of bisyllabic words seems to be a familiar pattern for Germans. Evidence comes from Friedrich (2003) who verified the assumption in a production experiment with German speakers. Participants heard an initial syllable, either derived from a stressed or unstressed word position (e.g. /bal/ from 'Balsam/balm or /ak/ from Ak'teur/active participant). They were asked to complete.

9 The definition follows Wiese (2000) who related this expression to proper, relatively formal present-day German speech (ibid., p. 2).
the word. Friedrich’s test persons produced more words with stress on the first syllable even when the played sequence came from a word with an unstressed first syllable. Some years before Friedrich, Jongenburger (1996) had already reported similar results for Dutch.

2.5.2. Perception

The functionality of stress as a \texttt{wb} marker is based on the assumption that stress is responsible for lexical activation. Our perception experiment will follow-up the question if stress violation hampers word recognition. We analyse mis-stressed words, e.g. \texttt{*Wasser\textsuperscript{wirtschaft}}/water economic (correct: \texttt{'Wasser\textsuperscript{wirtschaft}}) as well as unstressed words, e.g. sequence [díb@] taken from the utterance \texttt{die Belüftung der Wohnung/the ventilation of the flat}. In the last example, the German female definite article \texttt{die} and the prefix <\texttt{be}> - spoken in succession - were cut out of the given utterance. The resulting string had no full vowels, hence no stress, while the competing word \texttt{Diebe/thieves [díb@] has stress on the first syllable.}

Basically, the arrangement of strong and weak syllables in words plays an important role in the perception process. Wiese (2000) assumed that for \texttt{SG} stress is defined through a strong-weak relationship between syllables. Domahs et al. (2008) supported this hypothesis with their experimental results. They analysed to what extent mis-stressed words were tolerable and found that this depended on the metrical composition of the words. This is interesting because their results show that lexical activation of a correct candidate might work despite a wrong stress pattern. The authors monitored brain responses to correctly stressed trisyllabic words (e.g. \texttt{Vita\textsuperscript{min}}/vitamin, \texttt{Bi\textsuperscript{kini}/bikini, 'Ananas/pineapple}) and their incorrectly stressed forms. Targets were presented acoustically and visually. Event-related potentials (ERP, brain response evoked by cognitive/sensory processes) were measured. Of the incorrectly stressed forms, \texttt{*Ana\textsuperscript{nas}} was more tolerable than \texttt{*A\textsuperscript{nanas}}, \texttt{*Vitamin} was more tolerable than \texttt{*Vi\textsuperscript{tamin}}. Stress violations on the ante-penultimate syllables produced earlier effects than violations on the penultimate or the final syllable. The authors argued that it is “the position of an incorrectly stressed syllable rather than the position of a de-stressed syllable that determines the point at which participants detect a stress violation” (ibid., p. 27). If stress violation demanded a re-parsing of syllables into feet - \texttt{(Vi.ta/('min) into *Vi.(ta.min)} or \texttt{Bi.(ki.ni) into *(Bi.ki/(ni)} - test persons rejected the stress pattern immediately.
while stress violations that did not call for a metrical re-parsing procedure were better tolerated.

Friedrich (2003) also analysed the perception of German stress from a neurolinguistic perspective. She measured listeners’ EPRs while they heard incorrectly stressed words like *Am'boss/ambos or *'Abtei/abbey (correct: 'Amboss, Ab'tei). Incorrect stress generally delayed the test persons’ behavioral responses. She noticed, however, that only a stress violation of initially stressed words elicited a reliable ERP-effect. Other than Domahs et al. (2008), Friedrich (2003) examined bisyllabic nouns with the highly frequent strong-weak syllable pattern. The speech signal is thought to activate multiple entries stored in the listeners’ mental lexicon. Stress violation can play a role in the activation process, but other processes (e.g. lexical, semantic) may interfere to such an extent that stress becomes a minor factor. For her target *Am'boss, for instance, Friedrich assumed that the found EPR-effect might, amongst other possible factors, be influenced by lexical frequency. According to the CELEX database, first syllable /am/ is shared by approximately 110 German words 'Amsel/blackbird, 'Amrum/name of a German island, etc., most of them have initial stress.

Studies on stress perception were also carried out for various other languages. Cutler and Clifton (1984) showed for English that the acceptability of deliberately mis-stressed words depended on the position of the mis-stressed syllable. Listeners were asked to make lexical decision tasks about visually and acoustically presented targets. Cutler and Clifton found that recognition of English disyllabic words, which were incorrectly stressed on the final syllable (*ti'ger), was severely delayed (up to 200 ms delay in a semantic category decision task). On the other hand, the recognition of words with incorrect stress on the initial syllable (*typhoon) hardly suffered. English listeners, Cutler and Clifton explained, are familiar with “incorrect” stress on the initial syllable. This is because this kind of stress shift regularly occurs in spoken English, which is sometimes but not always connected to a change in lexical meaning, e.g. thir'teen, 'thirteen.

McQueen, Norris, and Cutler (1994) demonstrated in a word spotting experiment how stress influenced the activation of lexical competitors. The English word mess was detected more rapidly when it was embedded in nonsense word né'mes (no competitor) than in do'mes (competitor: do'mestic). McQueen, Norris, and Cutler (1994) also addressed metrical issues. Both do'mes and né'mes had a weak-strong (WS) syllable pattern. The authors tested activation levels compared to word pairs with a strong-weak (SW) syllable
pattern, while the first condition (competitor/no competitor) was kept. Target *sack* for instance was embedded (1) in a nonsense string *sackrek* (SW, no competitor) and (2) in a sequence *sakrif* which is the beginning of a meaningful word (SW, with competitor *sacrifice*). The result was that WS items were detected more quickly than SW items. WS sequences had error rates below 50 per cent. WS targets without lexical competitors (*nemes*) had the shortest reaction time. In the SW group, segmentation presented the greatest obstacle: In both example sequences, *sack* had been activated with the first syllable and with it several other competitors. Error rates for *sackrek* and *sakrif* were above 70 per cent. Astonishingly, *sakrif* as the sequence with the competitor (*sacrifice*) had a slightly lower error rate. Here again, the complexity of factors influencing the lexical activation process becomes apparent. The findings of McQueen, Norris, and Cutler (1994) are in line with the Metrical Segmentation Strategy (Cutler and Norris, 1988) which predicts that WS items have an advantage because lexical segmentation starts at the onset of the second (strong) syllable. Segmentation of SW items, on the other hand, starts right at the beginning of the string. Thus, when the target word and carrier word began at the same time, lexical competition was higher and the detection process was delayed.

How closely stress and word segmentation are connected, was also shown by Cutler and van Donselaar (2001) in a word identification experiment for Dutch. Listeners were able to tell if isolated syllables belonged to a stressed or an unstressed word (e.g. syllable *voor-* and targets *voornaam*/first name, *voor*naam/respectable). Cutler and van Donselaar (2001) found that *voornaam* is not a homophone for Dutch speakers. The sequence *voornaam* activated the lexical representation of *voornaam*, while *voor*naam activated *voor*naam, but they did not activate each other sufficiently enough to produce a repetition priming. In a later experiment, Cooper, Cutler, and Wales (2002) performed a similar perception experiment and compared the performance of native English-speaking and non-native (Dutch) listeners. Stress-manipulated English targets were presented, for instance *music*, which contained the first syllable [mɪʃ] from *museum* instead of the originally stressed syllable [mjuː]. The task was to judge if the targets were real English words. If stress cues are responsible for lexical activation, syllable [mɪʃ] in *music* and syllable [mjuː] in *museum* differ and the one syllable will not activate lexical competitors associated with the other (Cutler, 2005, p. 276). English native listeners performed poorer that the Dutch listeners. Tyler and Cutler (2009) explained that English listeners rather evaluate vowel quality than suprasegmental information, whereas Dutch listeners use suprasegmental cues like degrees of stress to a greater extent.
The tolerance of English listeners towards mis-stressed words had been shown earlier by Small, Simon, and Goldberg (1988). They argued that the recognition of mis-stressed words was possible as long as the resulting string was not a non-word and they substantiated this claim through a perception experiment. Homographs like ‘insert (noun)’ and ‘insert (verb),’ both meaningful, were embedded in sentences, whereas the noun had the stress-pattern of the verb and vice versa. Listeners were asked to press a button as soon as they heard the target. There was no delay in the speed of the test persons’ responses despite the noun/verb being mis-stressed. When the stress shift, however, produced a meaningless phonetic string (e.g. ‘che’mist’ instead of ‘chemist’), the reaction time was delayed.

2.6. Comparative vowel duration

The term “comparative vowel duration” describes the ability to disambiguate two sequences which are semantically/syntactically different but phonemically identical. Disambiguation is achieved through the temporal organisation of vowels in these sequences. As will be explained below, speech produced by native speakers of French show such an opposing vowel duration pattern and listeners are able to decode the information. Should such a strategy also work for German speech, we would have a parameter supporting wb detection in ambiguous sequences.

2.6.1. Production

Rietveld (1980) proposed a speaker strategy in French to distinguish phoneme sequences which had identical phonemic material, but differed in syntax, e.g. le muscat perdu/the spoiled muscat (wine) vs. le musc a perdu/the musk is evaporated. He recorded 10 French natives who uttered sequences in the style of the given noun phrase-sentence pair. Rietveld measured two vowels in each target sequence. In the noun phrase, these were the penultimate and the final vowel of the noun - in our example, /u/ and /a/ in muscat. In the sentence, these were the final vowel of the noun and the vowel of the auxiliary verb - in our example /u/ and /a/ in musc a. As a rule, one target was a single word, while the phonemically identical twin comprised two words and the wb between them. Rietveld found that muscat is spoken with short /u/ (95 ms) and long /a/ (120 ms), while musc a is pronounced with long /u/ (101 ms) and short /a/ (70 ms).
ms). He got similar results with other examples like le syndic sommé/the syndicate that has been called upon vs. le syndic a sommé/the syndic requested. Rietveld concluded that speakers applied a short-long vowel strategy for the noun phrases and a long-short vowel strategy for the phonemically identical sentences to disambiguate them. He also explained that the effect has to be evaluated in connection with the rhythmical structures of the investigated speech samples, see 2.6.2 for more details.

It has not been tested yet if such a mechanism also works for German, e.g. to distinguish phonemically identical word pairs like the compound Augenmaß (sense of proportion) and the noun-verb phrase [mit den] Augen maß (measured by the eye) by means of vowel durations.

2.6.2. Perception

Rietveld (1980) performed a perception test with 41 participants in order to corroborate his results obtained in the production experiment. Listeners heard the noun phrase and the corresponding sentence as pairs and had a booklet in front of them containing the pairs in the written form. For each pair, the participants had to mark if the noun phrase was heard first and the sentence second or vice versa. The participants mean discrimination score was 88 per cent. Rietveld concluded that the temporal organisation of an utterance supported discrimination.

Apart from vowel duration, there might have been more factors that enabled Rietveld’s listeners to distinguish between the two phrases. First, they heard both sequences in succession and were able to compare them directly. Second, and most importantly, the rhythmical structure of French gives important cues, which Rietveld pointed out himself. Word-final boundaries are signalled by a rise in fundamental frequency (e.g. Bahler, Coughlin, and Tremblay, 2011). In Rietveld’s pairs, these rises happen at different locations in the noun-phrase and the sentence. Rhythmical information might thus have been an important cue for the French listeners in the perception experiment.

The research of Christophe et al. (2003) also focused upon the connection between vowel duration and word segmentation. French native speakers participated in a word detection experiment. They heard a sentence with a local lexical ambiguity. Their task was to spot a previously given target word. The target word was either embedded across a prosodic wb or across a phonological phrase boundary. In the following example, phonological phrases are indicated through braces.
2. Theoretical background: Word boundary markers in German speech

Embedded word: chagrin

Prosodic wb
{Le livre} {racontait l’histoire} {d’un chat grincheux} {qui avait mordu} {un facteur}
The book told the story of a grumpy cat who had bitten a postman.

Phonological phrase boundary:
{D’après ma sœur} {le gros chat} {grimpait aux arbres}
According to my sister, the big cat climbed the trees.

Listeners reacted much faster in the phonological phrase boundary condition than in the prosodic wb condition. The vowel /a/ was lengthened at the phonological phrase boundary (phrase final lengthening) which might have given the crucial cue.

2.7. Geminates

A geminated consonant (from Latin “geminare” = doubling) stands in durational contrast to its singleton counterpart. This contrast is best shown using the model of McCarthy (1979) who introduced an autosegmental, phonological representation to illustrate the temporal organisation of speech sounds. For his systematisation, he listed the segments of a language and added a quantity tier in which long segments occupy two timing slots, and short ones occupy one. With his model, it was possible to have an abstract notion of quantity.

Usually, three forms of geminates are differentiated (Lahiri and Hankamer, 1988): tautomorphemic ones, concatenated ones, and those resulting from total assimilation. Their representations on the timing tier are the same. The principle is illustrated in Figure 2.2.

Geminates and their singleton counterparts differ in phonemic length (Reetz and Jongman, 2009). Many languages have contrastive length in their vowel inventory, while the contrast between long and short consonants is less common (ibid., p. 215). Consonant geminates can, for instances, be found in Italian, Finnish, Estonian, Bengali, Turkish or Swiss German.

Geminates are not part of the phonemic inventory of present day German. However, concatenated geminates can arise when one word ends and the following one starts with
The three forms of geminates (Lahiri and Hankamer, 1988) with examples from Tashlhiyt Berber (Ridouane, 2010) which is spoken in a part of Morocco. In a), word-initial stop /t/ occupies two timing slots (McCarthy, 1979) due to its duration. Word-final /t/ and word-initial /t/ become a concatenated geminate (also: contact geminate) in example b). Word-final stop /d/ and word-initial stop /t/ also result in a /t/-geminate, because /d/ undergoes a regressive assimilation process during which it loses the feature [- fortis] as shown in c).

![Figure 2.2](image)

Figure 2.2. The three forms of geminates (Lahiri and Hankamer, 1988) with examples from Tashlhiyt Berber (Ridouane, 2010) which is spoken in a part of Morocco. In a), word-initial stop /t/ occupies two timing slots (McCarthy, 1979) due to its duration. Word-final /t/ and word-initial /t/ become a concatenated geminate (also: contact geminate) in example b). Word-final stop /d/ and word-initial stop /t/ also result in a /t/-geminate, because /d/ undergoes a regressive assimilation process during which it loses the feature [- fortis] as shown in c).

the same consonant, like in *einem Mann*/a man, *dat*. The two phonemes at either side of the *wb* have contact, which is why we apply the term “contact geminate” to describe the resulting segment. Usually, the /m/ in the given example is produced longer than the corresponding singleton consonant as in *einem Schloss*/a palace, *dat*. (Mikuteit, 2007). Conceivably, this longer duration of /m/ in the case of a contact geminate could be used as a cue to *wb* detection. This suggestion was also put forward by Lea (1980) who analysed the duration of English sibilants in a machine-based recognition experiment. If a certain threshold in the sibilant’s duration was exceeded, he considered it as a candidate for division. In other words, segment duration can function as a *wb* marker.

For this thesis, we chose nasals to investigate the contrast between singletons and contact geminates in German. In German words, nasal singleton /n/ appears at word endings with heightened frequency, due to German morphology. Regular infinitive verb forms end in <en> (gehen/go) and so do the conjugated forms for the 1st and 3rd p. pl. (wir gehen, sie gehen/we go, they go). Strong verbs usually have a past participle that ends in <en> (gegangen/gone). Apart from that, feminine nouns ending in <e> take the allomorph <n> in the plural, for instance *die Rose*/the rose - *die Rosen*/the roses. In all these cases, German singleton /n/ is likely to form a contact geminate if the following word also starts with nasal /n/.
2. Theoretical background: Word boundary markers in German speech

2.7.1. Production

When we discuss singletons and geminates, we first have to look into their durational dimensions. How are they differentiated? There are a number of production studies on the durational contrast between singletons and geminates in languages with a consonantal length distinction. In the literature, the duration ratio between geminate and singleton nasals ranges from 1.61 to 2.87 (see Table 2.2).

<table>
<thead>
<tr>
<th>Language</th>
<th>Phonemes measured</th>
<th>Mean sing (ms)</th>
<th>Mean gem (ms)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madurese</td>
<td>nasals</td>
<td>90</td>
<td>145</td>
<td>1.61</td>
</tr>
<tr>
<td>Buginese</td>
<td>nasals</td>
<td>88</td>
<td>144</td>
<td>1.64</td>
</tr>
<tr>
<td>Swiss German (Thurgovia)</td>
<td>sonorants</td>
<td>62</td>
<td>114</td>
<td>1.85</td>
</tr>
<tr>
<td>Guinaang Bontok</td>
<td>/n/</td>
<td>73</td>
<td>140</td>
<td>1.92</td>
</tr>
<tr>
<td>Toba Batak</td>
<td>nasals</td>
<td>56</td>
<td>111</td>
<td>1.98</td>
</tr>
<tr>
<td>Japanese</td>
<td>/n/</td>
<td>68</td>
<td>178</td>
<td>2.62</td>
</tr>
<tr>
<td>Finnish</td>
<td>/n/</td>
<td>62</td>
<td>178</td>
<td>2.87</td>
</tr>
</tbody>
</table>

Table 2.2. Singleton-geminate ratios for different languages. Data come from Kraehennann (2003) for the Swiss German dialect of Thurgovia, from Aoyama and Reid (2006) for Guinaang Bontok, Aoyama (2001) for Finnish and Japanese and from Cohn, Ham, and Podesva (1999) for the three Austronesian languages of Indonesia: Buginese, Madurese and Toba Batak.

We are only aware of one study that investigated the durational properties of contact geminates in German. Mikuteit (2007) compared Bengali voiceless aspirated and unaspirated labial and dental geminates with German voiced and voiceless labial and alveolar stop geminates that have been derived by concatenation across morpheme boundaries. Her German stimuli were prefixed verbs, e.g. *mittesten*/*to include in a test*. Four native speakers participated in the production experiment. They had been instructed to utter words, which were visually presented on a computer screen, with a monotone (flat) intonation. Mikuteit found that “the closure of the geminates was on average 70 ms longer than that of the singletons” (ibid., p. 74). Her analysis gave a 1.82 geminate-singleton ratio. Kohler (1995) described this effect earlier as “Doppelkonsonanz” (double consonant area) that can only appear at morpheme boundaries.

Like German, French is a language without contrasting consonant length, but there are “secondary geminates”, as Meisenburg (2006) describes them. She investigated the following forms:
2. Theoretical background: Word boundary markers in German speech

1. Borrowings from Latin (irreël/unreal [iʁɛ̃l]) which were not investigated further
2. Sequences of identical consonants due to vowel deletion (je courrais/I shall run [kœ̃ʁɛ̃s] vs. je courais/I run [kœ̃ʁɛ])
3. Schwa deletion or non-emergence at morpheme boundaries (netteté/neatness [ⁿɛt̷ɛ])
4. Other position-induced geminates without schwa-intervention (pour Raymond/for Raymond [puʁœ̃rɛ̃mœ̃])
5. Geminate/singleon contrast (tu me mens/you lie to me [tym̷a] vs. tu mens/you lie [tym])

In a production experiment, which was camouflaged as a word finding and reading speed test, Meisenburg asked French speakers to produce sentences that contained targets with consonantal length constraints: Ça me frappe pas vraiment/This doesn’t really hit me and Il le frappa deux fois/He hit him twice. Her results showed that speakers did not produce consonantal quantity contrasts. Long consonants in French, Meisenburg concluded, are not relevant on the phonological level. Meisenburg’s study is of interest for our own production experiment, where speakers are similarly prompted to produce (contact) geminates.

An experiment conducted by Kabak, Reckziegel, and Braun (2011) followed-up the question if speakers of a language without phonemic geminates are able to produce such sounds. The study is not directly related to our own research but mentioned here, because it gives an insight into the learnability of geminate timing. Kabak, Reckziegel, and Braun (2011) described how German native speakers performed when learning a second language (L2), here Italian, which has true geminates in its phonetic inventory. Speakers without knowledge of Italian were compared with advanced learners, who had studied Italian between 5 and 10 years in high school and university and/or had lived in Italy between 6 months and 3 years. Native speakers of Italian formed the control group. The authors investigated whether L2 learners were able to produce a native-like timing of geminates. All participants heard a nonce word via headphones, spoken by a native Italian. Nonce words contained word-medial consonant geminates or singletons (/tupːa/, /tupa/). After hearing the target the second time, participants produced the test word themselves. Both groups – the one without knowledge of Italian and the one with advanced knowledge of Italian - were able to produce singletons and geminates with a durational contrast, despite the absence of such a contrast in their native language. The advanced learners’ group performed better, approaching the ratio of the native Italian
control group. For both German groups, geminate durations were significantly shorter, when compared to the control group.

**Long singletons**

In this section we summarise effects that lengthen singletons. The topic is of interest, because /n/ - one of the nasals we investigated in connection with contact geminates - appears frequently in word-final position and word-final segments are prone to lengthening. “Final lengthening reflects a general tendency to decelerate towards the end of a chunk because nothing else remains to be produced from the buffer” (McQueen and Cutler, 2010, p. 541). Effects of final lengthening were, for instance, reported by Salverda, Dahan, and McQueen (2003) who examined the temporal organisation of boundary segments. The authors found that monosyllabic words (ham) are longer than a phonemically identical syllable in a polysyllabic word (hamster). The examined language was Dutch. They used the word (hamster) as a carrier word and replaced the first syllable by the monosyllabic word (ham). Salverda, Dahan, and McQueen (2003) monitored the eye movement of listeners who heard the manipulated carrier word and simultaneously saw pictures of a ham, a hamster and two unrelated items. Test persons preferably looked at the picture representing the monosyllabic word. The experimental results suggest that there was an effect of segment final lengthening. The authors concluded that “the acoustic signal contained information that a purely phonemic description cannot capture” (ibid., p. 77). The monosyllabic word and the ambiguous syllable of the polysyllabic word differed with respect to temporal organisation.

Hofhuis, Gussenhoven, and Rietveld (1995) tested the effect of segmental lengthening at the end of five types of prosodic boundaries in Dutch with the prosodic wb as the lowest and the utterance boundary as the highest instance. Results showed that higher prosodic boundaries lead to more final lengthening than lower ones. Lengthening affected the vowel, as well as the consonant that followed it, but it was the consonant that showed stronger lengthening effects, even when it was a stop.

Singleton segments can also be lengthened in domain-initial position, a phenomenon referred to as phrase-initial strengthening. Nakatani (1978), for instance, found that word-initial consonants were longer than equivalent consonants in word-medial and word-final positions. He recorded reiterate speech consisting of “ma mama” phrases that mimicked typical sentence constituents. Afterwards, all durations of /m/ were measured. Apart
from word position, domain position is an important factor that influences singleton duration. Evidence for domain-initial strengthening, based as well on reiterate speech, was also reported by Fougeron and Keating \cite{1997}. They measured tongue-to-palate (linguopalatal) contact as an alternative method to express consonantal strength. Speakers produced equations, e.g. \textit{89 times (89 plus 89 plus 89) = alot} using the syllable /\textit{no}/, \textit{nonono no (nonono no nonono no nonono) = a lot}. According to their positions within the phrase, reiterate syllables were defined as initial, medial, or final. Linguopalatal contacts were recorded for each /\textit{n}/ in the respective position. Nasals in initial position in higher prosodic domains had more contact than initial nasals in lower domains. Onaka et al. \cite{2002} got similar data for Japanese. Prosodic position influenced the phonetic and acoustic properties of consonants. They examined /\textit{t}/ at five prosodic boundary levels: utterance, \textit{IP} accentual phrase, word, mora. The consonant was preceded and followed by the vowel /\textit{a}/ in all these positions. Like in the experiment of Fougeron and Keating \cite{1997}, linguopalatal contact was measured and, among other parameters, duration of the phoneme. More contact was found at higher prosodic domains than in lower ones. Also, the duration of /\textit{t}/ decreased from higher to lower domains.

A study of Cho, McQueen, and Cox \cite{2007} also investigated domain initial strengthening effects. They contrasted potentially ambiguous sequences such as \textit{bus tickets} (lexical competitor: \textit{bust}) coming from utterances with the target in different domain positions, e.g. (1) \textit{John forgot to buy the bus tickets} and (2) \textit{When you get on the bus, tickets should be shown to the driver}. While the word \textit{tickets} was in \textit{IP}-medial position in sentence (1), the same word started an \textit{IP} in sentence (2). The authors created spliced versions of their sentences. Here, the initial two phonemes of the second word (in our example /\textit{ti}/ in \textit{tickets}) were either taken from another utterance with the same prosodic structure or from an utterance equalling the other of the two prosodic structures tested. The sentences were played in a cross-modal priming experiment to listeners. While the recognition of target \textit{tickets} worked independent from the prosodic structure, listeners recognised the word \textit{bus} easier if the word onset of \textit{tickets} originated from an \textit{IP}-initial position. The researchers concluded that \textit{IP}-initial segments are articulated more strongly in terms of spatio-temporal expansion. \textit{IP}-initial segments also show greater coarticulatory resistance and were produced longer than word-initial consonants in lower domains, e.g. the \textit{wb}.

Turk and Shattuck-Hufnagel \cite{2007} demonstrated for American English that boundary related lengthening has to be seen with reference to the location of main stress. Thus,
a non-initial or a non-final syllable can also be affected. In words like *Carolina* or *Mississippi*, word stress is on the penultimate syllable which has a good chance to be lengthened.

Another interesting aspect is that lengthened singletons at prosodically important positions are likely to be followed by a silent interval. This might for instance be the case at the end of an [IP] at the end of a sentence constituent, or at the end of a multiword like a noun phrase (Levelt, 1993, p. 380). Such a break is “usually followed by a pause of more than 200 milliseconds” (Levelt, 1993, p. 385).

### 2.7.2. Perception

Gemination is also an interesting topic from the perception point of view. The question is: When is a phoneme recognised as geminated by listeners who have geminate phonemes in their language inventory? Hankamer, Lahiri, and Koreman (1989), for instance, showed in a perception experiment how Bengali and Turkish test persons, who have a singleton-geminate opposition in their respective native languages, categorised singletons and geminates. The presented stimuli were words with word-medial non-geminate stops, and the closure time of these stops was artificially lengthened in steps of 10 ms so that the sound eventually became a geminate. Both the word containing the non-geminate, as well as the one with the tautomorphemic geminate, were meaningful, e.g. Bengali *[pata]*/leaf vs. *[pat:a]*/whereabouts, Turkish *[eti]*/meat, acc. vs. *[et:i]*/do, past tense. Bengali listeners categorised voiceless stops as geminates above chance if closure duration exceeded 140 ms, while Turkish participants did so if closure duration exceeded 132 ms. In the cited study, listeners with singleton-geminate-opposition in their native language showed a fine-tuned language-related ability to distinguish between closure duration of a singleton stop and a geminate.

### 2.8. Interaction of word boundary markers

We have already mentioned that **wb** markers often co-occur. Word-initially, glottal stops and creaky voice may appear together (Wesener, 1999). After a silent interval, glottalisations of word-initial vowels are more frequent (Krech, 1968, Kohler, 1994). Word/phrase finally boundary-related segment lengthening was observed by Hofhuis,
Theoretical background: Word boundary markers in German speech

Gussenhoven, and Rietveld (1995). The higher the domain of the boundary, the more likely the additional use of a silent interval to mark the prosodic juncture becomes. A \textbf{wb} marker accumulation rather happens in prosodically higher domains like IP boundaries than in lower domains like word boundaries. In the case of an IP-final word, the domains fall together. A detailed study of the interaction between \textbf{wb} markers was not possible within the scope of this thesis. We will discuss the topic only briefly and exemplarily refer to the experiment carried out by de Pijper and Sanderman (1994) who have proposed a method to analyse \textbf{prosodic boundary strength (PBS)} in Dutch. The study had been cited earlier in connection with silent intervals. In general, the aim of the authors was to investigate if \textbf{PBS} is predictable as an independent variable. They asked three speakers (one professional, two non-professionals) to produce sentences in two ways, (1) in the relaxed, hesitation-free style of a newscaster, and (2) in their own, individual way; speakers were free to render their reading style. The recorded sentences were played to listeners who had to make judgements about where perceptual boundaries were located. They had been given a short introduction about prosodic boundaries (“A speaker clearly structures the speech by dividing it up into units, with audible boundaries between the units”). The judges were asked to score each \textbf{wb} in the utterance on a 10-point scale: “1” corresponded to a very weak boundary, “10” corresponded to a very strong one. As a first result, there was a high agreement among the participants of the perception test about the location of the boundaries. In a second step, the recorded material was analysed phonetically. Melodic cues, silent intervals, declination (pitch range) reset and to a limited extent pre-boundary lengthening, were the examined boundary markers. Their appearance was checked on those boundary locations perceived by the listeners; the allocated scores were considered as well. PBS was highest when a silent interval and a melodic cue were present. All three speakers used the same melodic cues, but the professional speaker performed pre-boundary vowel lengthening more distinctly and inserted more silent intervals.

2.9. Dialect specifics

Our perception experiment also touches upon the issue of dialect and how it influences \textbf{wb} recognition. The perception studies introduced in the following focus upon the question
2. Theoretical background: Word boundary markers in German speech

if and how dialects were distinguishable from standard pronunciation. One factor that characterises German dialects is that they differ in their intonation contours. This topic was addressed by Peters et al. (2002) who tested the perception of Hamburg and Berlin German. Listeners from both areas had to make judgements about intonation pitch contours that included their own local contour and non-local ones. They were asked to judge the local character of utterances by allocating points. The scale ranged from “sounds by no means like Hamburg German” (1 point) to “sounds very much like Hamburg German” (7 points). Participants who were familiar with both the local variety and some other non-local variety performed better than listeners familiar with only the local variety. The authors concluded that a successful performance of the test persons did not only depend on true recognition of the local contours, but that listeners also used an elimination strategy that enabled them to identify non-local contours correctly. In this elimination process, they apparently excluded “those contours that they were most familiar with” (ibid., p. 133) and recognition worked according to the scheme: not familiar=non-local.

The topic of dialect identification was also addressed by Clopper and Pisoni (2005). In evaluating a variety of perception studies, the authors explained: “Naive listeners can make reliable judgments about where an unfamiliar talker is from without explicit instructions about what to listen for” (ibid., p. 327). Another conclusion drawn from their evaluation was that the performance of the test persons increased with growing experience and greater exposure to multiple dialects. This is an interesting point for our own study on dialect specifics which includes a dialect speaker and listeners with dialects.

Wb markers (e.g. stop allophones) are affected by dialects. The Saxonian dialect, for instance, is characterised by a lacking distinction between [+ fortis] and [- fortis] stops (Kohler, 1995). The question that arises from Clopper and Pisoni’s conclusion is whether wb recognition works independent from speakers’ and listeners’ dialects.

In this chapter, wb markers were introduced from the production and perception point of view. The interaction between some wb markers was briefly described and the influence of the factor dialect on speech perception was mentioned. The following chapter 3 starts the experimental part of this thesis. A corpus of read speech was created that contained target words with wb markers. This was achieved through the merging of neighbouring words or word parts to one new word. The production data were investigated with
2. Theoretical background: Word boundary markers in German speech

the aim to get a first idea about the specifics of the wb markers under discussion, e.g. their place and frequency of occurrence. Results were compared to the findings of other researchers as discussed previously in this chapter.
3. Production experiment

3.1. Introduction

The production experiment was designed to study unpredictable wb markers. These were glottal stops and creaky voice, stop allophones, silent intervals, comparative vowel duration and contact geminates. The term “unpredictable” is used because the application of these wb markers is not certain; speakers might use as well as omit them.

In order to study wb areas, a special corpus of speech data had to be produced. For that purpose, minimal pairs with and without wb areas were elicited. The sequence Monat, for instance, can also be found in Mona Tölle/first name and family name. By merging Mona and the next word’s initial stop /t/, an artificial word was created containing a “wrong” final stop allophone. The word pair Monat - Mona T allowed us to investigate the stop allophone /t/ to clarify its potential as a wb marker.

Following this principle, a list of word pairs was created in which word pair constituents were phonemically identical, but the merged constituent had the potential to differ in at least one phonetic feature caused by the wb that was included, see also Table 3.1.

The embedding of words in other words is a frequent phenomenon in many languages (cf. section 1.2, Luce, 1986). Target-oriented embedding appears to be a useful method for word recognition tasks. Usually, researchers have used this technique for the production of acoustic stimuli in preparation of perception experiments. Some experiments with embedded words were introduced in chapter 2 McQueen, Norris, and Cutler (1994), for instance, embedded real English words in non-words (mess in domes) with the purpose of studying segmentation strategies. For French, Christophe et al. (2003) compared the

---

1Note that sequences like Mona T were simply cut out of the audio stream and analysed without further manipulation. The term “merger” alludes to the fact that the sequence is not a natural word but a constructed one.
perceptibility of targets that were embedded across a prosodic or across a phonological phrase boundary. Salverda, Dahan, and McQueen examined Dutch speech and exchanged the first syllable of a carrier word (e.g. hamster) by an ambiguous monosyllabic word (ham).

For our own production experiment, speakers were recorded reading texts. Word pair constituents were cut out of their respective sentences and analysed. Our special attention was directed towards those mergers that had received markers. Depending on the markers, we examined different characteristics: (1) place and frequency of occurrence (glottal stops/ creaky voice, silent intervals), (2) differences in intensity levels (stop allophones), (3) duration (silent intervals, comparative vowel duration, contact geminates). More generally we examined if markers were applied at all and/or in what quality on quantity. Remember that the markers /l/-allophone and stress were not analysed because their application is regulated.

Table 3.1. Unpredictable markers with word pair examples.

<table>
<thead>
<tr>
<th>Word boundary marker</th>
<th>Natural word</th>
<th>Merger</th>
<th>Why unpredictable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal stop/creaky voice</td>
<td>mager/skinny</td>
<td>mag/er/‘likes he’</td>
<td>Place and frequency of occurrence</td>
</tr>
<tr>
<td>Stop allophone: initial stop in final position</td>
<td>Monat/month</td>
<td>Mona_Tölle/first and family name</td>
<td>difference in aspiration/intensity levels</td>
</tr>
<tr>
<td>Stop allophone: final stop in initial position</td>
<td>klebt/to stick, 3rd p. pl.</td>
<td>[das Handwer]k_lebt/craftsmanship is alive</td>
<td></td>
</tr>
<tr>
<td>Stop allophone: medial stop in initial position</td>
<td>Palau/island in the West Pacific</td>
<td>[ihren Grappa]lau/warm zu trinken/to drink their Grappa lukewarm</td>
<td></td>
</tr>
<tr>
<td>Stop allophone: initial stop in medial position</td>
<td>Skalp/scalp</td>
<td>[erste]s_Kalb/first calf (stop /b/ is uttered as /p/ due to German terminal devoicing)</td>
<td></td>
</tr>
<tr>
<td>Silent interval</td>
<td>Guthaben/credit balance</td>
<td>gut_haben/being well off</td>
<td>Place and frequency of occurrence/duration</td>
</tr>
<tr>
<td>comparative vowel duration</td>
<td>Augenmaß/sense of proportion</td>
<td>[mit den] Augenmaß/ measured by the eye</td>
<td>duration</td>
</tr>
<tr>
<td>Contact geminates:</td>
<td>einem Mann/one man, dat. - einem Schloss/one castle, dat.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2One might argue that stress distribution is not necessarily regulated. Dialect speakers from Baden
The production experiment that follows is based on the recordings of three texts, see Table 3.2. In the case of the contact geminates, however, more speech material was required which lead to additional recordings of a fourth text which is described separately (section 3.4.6). A summary of all results obtained in the production experiment concludes the chapter.

3.2. Methods

For the statistical evaluations, linear mixed-effect modelling was used, which had the advantage that all speakers independent from their speech tempo could be included in the analysis as random factor.

3.2.1. Participants

The speaker pool comprised 19 speakers - 8 female and 11 male participants. For various reasons, there was a different number of participants for each text to be read, see Table 3.2 and section 3.2.3. Speakers were between 24 and 60 years old and came from Baden-Württemberg, Saxony, Saxony-Anhalt, Lower Saxony and Mecklenburg-West Pomerania. Nine persons were dialect speakers (Saxon, Mansfelder Land dialect [Saxony-Anhalt], Swabian, Baden dialect), all others spoke SG as defined by Wiese, 2000, p. 2 (proper, relatively formal present-day German speech). The participants classified as speakers of SG spoke proper contemporary German and had no apparent dialect. However, their speech was not entirely free from casual or regional colourings.

Speakers were academics and non-academics, but none of them had a linguistic background. They were unaware of the aim of the research.

or Swabia, for instance, frequently utter words with non-standard stress in spontaneous speech (e.g. Büro, SG: Bü’ro). However, we did not expect to find non-standard stress in speech material generated in a reading task which usually involves a more careful pronunciation. This assumption proved true later; readers uttered all words with standard stress.

Participation was voluntary and unpaid. Speakers often had a fixed time frame for their participation which was occasionally too short to record all texts.

We agree with Wiese who observed that it is “debatable whether there is an actual pronunciation which is totally purified of all regional connotations” (ibid., p. 2).
3. Production experiment

<table>
<thead>
<tr>
<th>Text number</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Newspaper article</td>
<td>Short story</td>
<td>Contextually unrelated sentences/mini episodes</td>
</tr>
<tr>
<td>Number of readers</td>
<td>19</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>Number of dialect speakers among readers</td>
<td>9</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Number of sentences</td>
<td>23</td>
<td>27</td>
<td>59</td>
</tr>
<tr>
<td>Contiguous in context</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Number of word pairs</td>
<td>7</td>
<td>8</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 3.2. Reading texts and readers

3.2.2. Materials

Reading texts were created that aimed to produce word pairs consisting of a natural word and a merged twin that carried at least one wb marker. Several mergers had the potential to receive more than one boundary marker. B_all, for instance, which was derived from the sequence ab_Allensbach/from Allensbach (small but well-known town in the South of Germany), started with final stop /b/ in initial position. Furthermore, the vowel /a/ might be uttered with a glottal stop and/or creaky voice, since it comes from a word-initial position. Furthermore, a silent interval between /b/ and /a/ is possible. Contrary to this, the target words chosen with the purpose of studying comparative vowel duration had little potential to receive any markers at the wb. We were interested in receiving nearly identical sequences that enabled us to concentrate solely on the analysis of the durational patterns of stressed vowels.

In the reading texts, the natural word and the sequence out of which its twin was to be derived did not appear in the same or in neighbouring sentences. In total, 34 word pairs were created. Nasal contact geminate and singleton targets were not contrasted as word pairs. Contact geminates arose when one word ended and the neighbouring word started with a nasal. All geminated segments that appeared in the three reading texts were compared to all nasal singleton items stemming from all target words of the same texts. The experimental layout for geminates and singletons is explained in more detail in section 3.4.5.

There are two exceptions: Skonto/allowance and /[eigene]s_Konto/one’s own bank account, as well as tankt and Tank, t, appeared in one sentence. All reading texts can be found in the Appendix section A.1.
In order to study stop allophones, the reading texts contained sequences which licenced different degrees of aspiration.

1. Initial segments uttered with high aspiration.

2. Fricative-stop clusters, nasal-stop clusters, and lateral-stop clusters in which the stop had little or no aspiration.

3. Intermediate forms like word-final and medial stops.

### 3.2.3. Procedure

Participants were recorded reading texts. The first text was written in the style of a newspaper article, the second one was a short story, and the third text consisted of unrelated sentences (for more information, see Table 3.2). Remember that not all participants read all texts - Table 3.2 also states the number of readers per text. All 15 readers who read the second text had also read the first text. Among this group of 15 readers were three readers who also read the third text. That leaves four participants who only read one of the texts (which was the first text). All participants were recorded individually. Speakers who read more than one text, were usually recorded in one session.

Prior to the actual recording, the reading in front of the microphone was practiced. The aim was to make the speakers acquainted with an unfamiliar situation. This training period was also used to adjust the recording equipment to the voice of the respective speaker. Recordings were either made in an anechoic chamber at the University of Applied Sciences Konstanz or in quiet private rooms using a Sony ECM MS 957 microphone and an Edirol R-1 Portable Digital Recorder (24-bit/44.1 kHz). In the anechoic chamber, printed texts were read; in home surroundings, speakers read either from paper or from a computer monitor.

Each of the participants was instructed to read the text aloud. The experimenter was present during the recordings. If a speaker mis-spoke a target word, the experimenter took a note and asked the speaker later to repeat the paragraph in which the error had occurred. Participants were not interrupted during the recording, the process of error correction was carried out after one text had been recorded in full.

Target words from all recordings were cut, segmented and labelled with Praat (Boersma and Weenink, 2005), resulting in a text file for every target word. In the labelling task,
we had to address the problem that speech sounds are not sharply divided from each other. When speech is produced, the articulatory organs - lips, tongue and jaw - cannot move freely and abruptly, but do so in a highly co-operative manner. Adjacent phonetic segments interact with each other (Kashino, 2006). The transition from one phonetic segment to the next is summarised under the term “coarticulation” (ibid.). Consequently, “the vocal tract configuration at any point in time is influenced by more than one segment” (Farnetani, 1999, p. 316). Our decision, where to place a segment boundary, was motivated by four sources of information: spectrogram, fundamental frequency, formant movements and auditory impression. Spectrograms support the identification of voiced and unvoiced phonemes. A range of phonemes have typical representations in the spectrogram (e.g. vowels), while others are more difficult to determine (e.g. /l/). Fundamental frequency as the lowest frequency of a complex tone gives important cues about the location of sonorant segments (Reetz, 2003). Formant movements provide valuable information about segment transitions caused by tongue height, tongue position and lip rounding. Auditory impression was a useful verification method for the labelled units, especially for those with irregularities in the speech signal (e.g. creaky voice sections) or adjacent phonemes with few differences in the signal patterns (e.g. sonorants and vowels).

For the labelling task, we adopted criteria proposed by Turk, Nakai, and Sugahara (2006), as listed below. Some of their criteria were adjusted and/or extended for the purpose of our research.

Labelling criteria were as follows:

1. Vowels: A label defines the interval duration between a consonantal release landmark and a consonantal constriction landmark in CVC clusters. The segment boundary of a phrase-initial vowel was placed at the voice onset time.

2. Stops: Closure and burst were labelled separately in order to perform analyses that concentrated solely on the consonantal burst (measurement of burst intensity in stop allophones). For velar stops, which are often accompanied by multiple bursts, the last glottal pulse peak was chosen to mark the offset of the stop (Turk, Nakai, and Sugahara, 2006, p. 7).

3. Glottal stops: These sounds were usually identifiable through a period of silence followed by a burst and some irregular pitch periods which gradually became regular as the phonation of the vowel set in. (Alternatively, the burst was followed by a
creaky segment as described below.) In our target words, glottal stops occurred at predictable positions (word-initially before vowels, morpheme-initially as stop substitutes), which supported their location. Auditory verification was carried out for all visually identified glottal stop segments and it was the determining factor when questionable segments had to be evaluated.

4. Creaky voice: The criteria that pointed to creaky voice were irregular periods (e.g. jitters) and - when present - a low F0. All segments were also evaluated acoustically, which - like with glottal stops - finally decided about debatable cases. Creaky voice was included in the label for the segment that had been uttered creaky, e.g. [a (cv)].

5. Silent intervals: The labelling criterion was a flatline/near flatline in the signal indicating silence or near silence (acoustically verified), the used label was “0”. Stop closures, for instance, were labelled as silent intervals.

6. Nasals: The abrupt changes in the spectrogram which are typical for nasals were used to define segment boundaries. However, transitions between nasals and vowels were the sections which proved to be difficult. Segment boundaries were occasionally obscured. Turk, Nakai, and Sugahara (2006) suggested that such instances can be dealt with in two ways. First, sections with uncertain boundaries can be annotated as “questionable”. This possibility was not considered, since nasals played an important role in our study of the nasal singleton-geminate contrast. Instead, we followed the authors’ second suggestion to segment according to a chosen policy: place a boundary either earlier or later, when in doubt, and to do this consistently (ibid., p. 16). We decided to place the boundary later in the nasal. This policy was not only applied to nasal-vowel transitions, but to all questionable segments of the dataset.

7. Lateral /l/: If possible, spectral discontinuities at the constriction onset and release (ibid., p. 15) were checked. However, transitions between /l/ and neighbouring phonemes were usually blurry, which made segment boundary placement rather difficult. In case of doubt, the segment boundary was placed later, as described above.

^Non-glottalised word initial vowels (e.g. breathy or soft onsets) were also identified but not distinguished; they shared one label.
8. Contact geminates: A contact geminate was defined when two nasals met at a [wb ] and the waveform showed one uninterrupted nasal. If there was an interruption between the neighbouring nasals – either a short period of silence or the signal showed a new onset for the second nasal despite the tonal connection of both phonemes – they were regarded as two singleton consonants. The principle is demonstrated in figure 3.1.

![Figure 3.1](image1.png)

**Figure 3.1.** In the sample *einem Monat*, the two nasals /m/ framing the wb were not uttered as a geminate, but as two singletons separated by a silent interval. In the other sample (*dem Mann*), a contact geminate was formed.

In the case of stop allophones, the measurement of aspiration (energy/duration) as a cue was not considered an appropriate method. The reasons for this decision were that 1) recordings had been made at different locations under very similar but not equal conditions and 2) speakers’ distance to the microphone had not been controlled. Raw values were found to be difficult to compare, a cross-speaker comparison of allophones was therefore avoided. Instead, we decided to use a more reliable cue, namely intensity levels of
stops as indicators for consonantal strength, and to perform a within-subjects/within-items analysis. The approach seemed reasonable, because the comparison of segments within a sequence would minimise distortions caused by a speaker’s distance to the microphone. The allophone’s intensity value was subtracted from the intensity value of the stressed vowel nearest to the allophone. Vowels represent the segment with the highest and most concentrated energy in a word (Jakobson and Halle, 2002). Consequently, when the difference between the stressed vowel and the stop allophone was little, the stop allophone had high intensity values approaching that of the vowel. If the difference was large, the stop allophone had much lower energy.

In total, 562 items out of the recorded 570 were suitable to be analysed. Eight items had to be excluded, because the required segments were not present. These were mostly targets with a missing or unreleased final stop. For each [\textit{wb}] marker under observation, all text files containing this marker were selected using a software written in C++. Data was written into a csv-file which could be imported into a spreadsheet for analysis.

The following section explains, which hypotheses were put forward and motivates the research questions that were followed-up.

3.3. Hypotheses and research questions

3.3.1. Glottal stops/creaky voice

We hypothesised that the glottal stop would prove to be a frequent and reliable \textit{wb} marker. Corpus studies by Krech (1968) and Kohler (1994) substantiated the fact that glottal stops are crucial border signals which are maintained independent of speech rate. Furthermore, we expected that creaky voice would be a regular companion of glottal stops as observed by Wesener (1999). He reported that glottal stops usually appear in conjunction with creaky voice.

Research questions were:

1. In what word positions do glottal stops and creaky voice segments appear?

\footnote{Contact geminates and singletons are not included here. They were tested for a contrast in duration and we therefore examined them separately, see section 3.3.3.}

\footnote{We thank Silvio Robel.}
2. How often are vowel-initial target words preceded by a glottal stop and how often not?

3. How often do glottal stops and creaky voice appear together?

3.3.2. Stop allophones

We expected to find different energy values in stop allophones depending on their position in a word, because our targets comprised both highly aspirated (high energy) initial segments as well as items without aspiration or with low aspiration (no energy/low energy), for instance fricative-stop clusters.

We hypothesised that intensity values of stop allophones vary according to their position in a word.

3.3.3. Silent intervals

We assumed that silent intervals would be longer at the wb of a merged word than silent intervals at the corresponding area of the natural counterpart. Silent intervals in natural words were regarded as purely articulatory-induced, while in some mergers deliberately inserted silent intervals would show up. We also hypothesised that in mergers with stops after the wb (Mona_Tölle, natural word: Monat/month), deliberately inserted silent intervals and stop closures would be hard to distinguish from one another. The latter can reach considerable maximum values (see Table 2.1). This will make a differentiation between the two types rather difficult.

Research questions:

1. What are the general durational dimensions of silent intervals between words?

2. Do mergers have longer silent intervals than natural words at the wb area in question?

3.3.4. Comparative vowel duration

We hypothesised that comparative vowel duration as proposed by Rietveld (1980) for French would also appear in German. The targets he contrasted, e.g. sequence [myska]
3. Production experiment

in *le muscat perdu/the spoiled muscat [wine]* vs. *le musc a perdu/the musk is evaporated*, had differing vowel durations depending on the syntactic structure of the origin. Targets from noun phrases (*muscat*) showed a short-long pattern for the measured vowels (here: /u/ and /a/), while targets from the sentence constructions (*musc a*) had been uttered with a long-short pattern. Our German pairs also originated from different syntactic structures. *Augenmaß/sense of proportion*, for instance, is a compound consisting of two nouns, the merged twin /mit den* Augen-maß/measured by the eye is a noun-verb construction. However, during the annotation of the data, a first impression about the potential of the German pairs could be obtained. This impression suggested that the word pair constituents would not contrast with regard to comparative vowel duration.

The research question was whether German speakers use vowel duration as a **wb** marker to disambiguate phonemically identical word pairs.

### 3.3.5. Contact geminates

**First test**

We hypothesised that in German, contact geminates might function as **wb** markers. Nasals were chosen to investigate this assumption. Nasal /n/ is special because it frequently appears at word endings (see section 2.7.1) and is thus well suited to form a contact geminate with a following nasal, but is also prone to be lengthened as a singleton in final position.

The language has no spoken nasal geminates within words. However, the concatenation of two homorganic nasals at the end of one and the beginning of the following word (*e.g. einem Mann/one man, dat.*) often results in a long nasal.

The goal of the production experiment was to investigate if the temporal organisation of German nasals differs across conditions. It was assumed that two homorganic nasals at the end of one and the beginning of the following word will be articulated as a geminate and that this geminate will have a significantly longer duration than corresponding nasal singletons. Hence, nasal duration could be used as a cue for **wb** detection. In other words, if the duration of a nasal exceeded a certain duration, there are in fact two nasals present framing a **wb**. Quasi-geminate /nm/ was included in the analysis of contact geminates. The temporal behavior of /nm/ was expected to be no different from the two homorganic geminates /nn/ and /mm/. We were aware of a competitive situation
between lengthened singletons located at [IP] boundaries and contact geminates, which might have similar durations.

The literature discussed in section 2.7 provides singleton-geminate ratios for nasals or sonorants in general for different languages. These ratios ranged between 1.61 and 2.87. It was expected that our own acoustic data for nasals would provide a ratio within the range of the cited values.

Accordingly, research questions were:

1. Is there a significant difference in duration between nasal singletons and contact geminates that recommend contact geminates as a suitable [wb] marker?
2. Is there a competitive situation between word-final lengthened singletons and contact geminates?
3. To what extent do durations of homorganic geminates /nn/ and /mm/ and quasi-geminates /nm/ and /mn/ differ?
4. What is the singleton-geminate-ratio?

Second test
A second test was designed to study singleton and geminate durations under more controlled conditions. Contact geminates analysed in the first test had not deliberately been placed in the three reading texts. The geminates were detected later during labelling. Consequently, the number of items was not equally distributed across conditions. However, the results of the first test are still worth considering, because the speech material has been produced by speakers without linguistic backgrounds. The second test, which followed up the temporal organisation of singletons and geminates in a more controlled way, used speech material produced by students of linguistic. The research questions remained the same.

3.4. Results and discussion

3.4.1. Glottal stops/creaky voice

The aim of this analysis was to examine the positions in which glottal stops and creaky voice occurred as single markers and as marker combination in general and at word-initial
3. Production experiment

vowels especially.

Glottal stops and creaky voice in all target words and all occurring positions had been labelled. As a rule, all target words with a word-initial vowel had received a label for the manner of the initial vowel’s articulation. Labels were allocated for glottal stop, for creaky voice and for the glottal stop-creaky voice combination. Apart from the mentioned glottal gestures, other forms were observed, too (e.g. breathy, soft onsets). Those items received the label “other forms”. Out of 150 targets with word-initial vowels, 104 were realised as with a glottal gesture, which was either a glottal stop, creaky voice, or a combination of glottal stop and creaky voice (see Figure 3.2). The most common form of vowel-initial glottalisation was creaky voice. Glottal stops were produced less often than creaky voice. This result is in agreement with Wesener (1999). However, his finding that a glottal stop occurred predominantly in combination with creaky voice, could not be replicated in this production experiment. Only one third of the glottal stops were accompanied by creaky voice articulation. We can only speculate about the reason for this. There was, however, one conspicuousity that caught our attention: Speakers of SG produced initial vowels with glottal stops slightly more frequently than dialect speakers (37 glottal stop items in total, including sole glottal stops and glottal stop-creaky voice combinations; 21 glottal stops produced by SG speakers, 16 by dialect speakers). For example, target *mag er/likes he* was occasionally uttered as *[maːɡə]* or *[maːɡə]* by dialect speakers. Such assimilated forms were especially observed for Swabian and Saxon speakers. While dialect speakers also produced glottalised forms like *[maːk ʔɛːə]*, *[maːk ɛːə]*, *[maːk ʔɛːə]* etc., we noticed a trend that speakers of SG did

Figure 3.2. Articulation of word-initial vowels in 150 examined segments.

Beside glottalised initial vowels, 12 targets with initial consonants [v] or [ʃ] preceded by a glottal stop were obtained, but will not be discussed further. Consonant-related glottalisation was also reported by Kohler (1999) who examined German plosives.

---

9Beside glottalised initial vowels, 12 targets with initial consonants [v] or [ʃ] preceded by a glottal stop were obtained, but will not be discussed further. Consonant-related glottalisation was also reported by Kohler (1999) who examined German plosives.
so more often. Creaky voice as a single \textit{wb} marker, however, was equally distributed. The production of fewer glottal stops in total leads in consequence to a smaller number of possible glottal stop-creaky voice combinations. Wesener’s speech material source was the Kiel Corpus. The majority of the speakers contributing to the Kiel Corpus stemmed from those regions of Germany where SG is spoken (Hamburg, Lower Saxony, Schleswig-Holstein). On the other hand, nearly half of the participants of our production experiment were dialect speakers. If our assumption is correct, the local background of the speakers who contributed to the respective experiments might have been a major influencing factor. The number of speakers and items analysed in our own experiment is, however, too small to generalise. A more comprehensive study is needed.

Forty-six vowel onsets were realised as non-glottalised. These targets showed either soft or breathy onsets or were assimilated forms, as explained above.

Word-medial (morpheme-initial) glottal stops appeared in 46 out of 131 potential targets that showed the typical sequence stop-schwa-nasal like target word \textit{Klappen}. A glottalised realisation [klaʔm] (here: glottalisation plus place assimilation of /n/ to /m/) was quite common. Apart from the mentioned phoneme sequence, there were no other appearances for word-medial glottal stops.

In section 2.1 we expressed our concern that creaky voice might be unreliable as an independent \textit{wb} marker. Non-boundary related creaky voice articulation appeared in 21 targets out of 67. Non-boundary related creaky voice articulation means that vowels (also: twice a nasal, and once a liquid) showed a creaky voice pattern somewhere in the word, but not at a boundary. We found non-boundary creaky voice, for instance, on /ˈö/ in \textit{Störenfried/trouble maker}. Twelve of the reported 21 targets were produced by one male participant alone. He articulated a great number of the target words with creaky voice, because he spoke with a low-frequency F0. Irregularities in the speech signal caused by slack vocal folds that do not fully vibrate, are one characteristic for speakers with a deep bass line.

Apart from word-initial creaky voice, we also found two word-final instances. Infante Rios and Perez Sanz (2011) differentiated between creaky voice and creaks (pulses) with the help of an EGG analysis and found that the two variants appear at different locations. Creaky voice is found utterance-medially, while a creak signals the end of an utterance and would, in consequence, be as much a \textit{wb} signal as word-initial creaky voice. In our data, the label “creaky voice” comprised both creaky voice and creaks. It is therefore
uncertain whether the two [IP] final creaky voice items that showed up were in fact creaks as defined by Infante Rios and Perez Sanz.

Speakers differed greatly in their laryngeal gestures for vowel-initial targets. There was no such thing as a “typical” pattern. This observation was also made by Kießling et al. (1995), who analysed laryngealizations in German speech in order to fine-tune a speech recogniser with regard to laryngeal gestures. Every speaker in their production experiment had different articulatory strategies with respect to the studied feature. Not one of the detected laryngeal properties was transferable from one speaker to another. A similar report came from Dilley, Shattuck-Hufnagel, and Ostendorf (1996) for American English.

3.4.2. Stop allophones

The aim of the analysis of stop allophones was to examine if their intensity levels relative to the next stressed vowel differed according to word position. Four speakers (three speakers of SG, one dialect speaker from Saxony) were analysed. Their speech material was later chosen as acoustic stimuli for the perception experiment, because they had uttered all word pair constituents in the best quality. In the case of the stop allophones, the speech material produced by these four speakers showed the strongest audible contrast, e.g. strongly aspirated initial stops versus unaspirated final stops in the case of the SG speakers. In the case of the participant from Saxony, syllable-initial stops showed little aspiration in general. However, his word pairs contrasted nevertheless, because many of his syllable-final stops showed interesting assimilation effects across the wb, which resulted in a well-audible allophonic difference. Altogether, we expected that if there was an effect of boundary condition, it would show up for the chosen four speakers.

Our analysis comprised of 21 word pairs with stop allophones. The intensity measurement of each stop allophone started at the burst, not at the closure point. Stops had been labelled accordingly. We used a Praat script that calculated mean intensity in dB across a labelled segment. This was carried out for both the stop burst and the nearest stressed vowel, then we calculated the intensity ratio. Here, the intensity value of the stop was subtracted from the intensity value of the vowel. For target Zug/train for instance, the average intensity values for the final stop /k/ and vowel /u/ were determined. If, for example, /k/ had a value of 32 dB and /u/ had a value of 41 dB, the

\[\text{Intensity ratio} = \frac{\text{Intensity of stop}}{\text{Intensity of vowel}}\]

We thank Henning Reetz for providing the script.
difference was 9 dB. The same procedure was done for the merged twin *zu_k/omm[n]/to come*. In the following analysis, vowel-stop differences in natural and merged words were compared.

We used linear mixed-effect modelling (Baayen, Davidson, and Bates, 2008) with vowel-stop intensity difference as dependent variable, boundary condition (natural/merged) as independent variable and participants and items as crossed random factors. We found no effect for boundary condition (p > 0.9, mean vowel-stop difference: 10.84 dB for merged items and 10.98 dB for natural items), see Figure 3.3. Standard deviation was high in both conditions (3.9 dB).

![Intensity values of stop almorphes](image)

**Figure 3.3.** Allophone intensity values calculated as vowel-stop ratios (dB value vowel minus dB value stop) in merged and natural items.

The experiment was therefore repeated with a second dataset. Data came from an additional recording that was carried out in connection with the perception experiment described in chapter 4.3.1. The second test with 22 stop allophone word pairs produced

Quite recently, a new method for mixed-effects data analysis was introduced by Cunnings (2012). He suggests to include a random adjustment for intercept and slope in order to account for the individual performance of participants on one side and a varying impact of items on the other side. Cunning’s publication was submitted at the time of writing this thesis. Exemplarily, we applied the proposed model to some of our calculations. The newly obtained results did not differ significantly from those calculated previously. In some cases, however, the model showed converging problems, either caused by an insufficient number of data or the structure of data.
by one speaker of SG neither showed an effect ($p > 0.2$, STD 4.1 dB, mean vowel-stop difference for merged items: 11.46 dB and for natural items: 13.18 dB). Our results from both calculations suggest that allophone intensity is independent from the allophone’s position in a word. However, both datasets showed a high standard deviation. A direct comparison of our results with the literature regarding the behavior of German stop allophones cited in section 2.2 was not possible, because the literature relates to aspirational cues.

3.4.3. Silent intervals

The criteria for the label “silent interval” was a flatline in the signal indicating silence or near silence. All of these sections had been auditorily verified. For mergers, the $\text{wb}$ area was observed, and when a silent interval was present, it was labelled. Natural words were examined in the same way. The area in the natural word corresponding to the $\text{wb}$ area of the respective merger was checked for a silent interval which when present was labelled as well (e.g. *Kodak*/name of a company - *Koda#k*[lingt]*/coda sounds, analysed areas are underlined). Silent intervals at other locations were not considered.

The analysis is based on 112 word pairs. These are the word pairs with both constituents having a silent interval at the $\text{wb}$ area (mergers) and the corresponding area (naturals). Not included in the dataset were 19 word pairs, in which one constituent had a silent interval and the other one had none. In 13 cases, the merger had a silent interval at the $\text{wb}$ area, and the natural word had none. In the other 6 cases, the situation was reversed. These word pairs were excluded from the analysis, because a “mean” duration of 0 ms would have distorted the statistical calculation.

Results were evaluated with linear mixed-effect modelling with silent interval duration as dependent variable, boundary condition (natural/merged) as independent variable and participants and items as crossed random factors. Outliers were removed. The comparison of silent interval durations in merged and natural targets delivered no effect ($p > 0.2$) even though silent intervals in mergers had a longer mean duration (76 ms) compared to natural words (60 ms). The obtained silent intervals had durations similar to stop closure times, so we cannot exclude the possibility that some of the analysed silent intervals in mergers with a stop after the $\text{wb}$ simply represent stop closures. Compared

---

12 In this calculation, outliers had not been removed due to the limited number of speakers and items.
13 For this calculation and others that follow using linear mixed-effect modelling, all residuals that fell outside the +/-2.5 sigma range were defined as outliers.
to the average stop closure durations in German speech based on the Kiel Corpus of Spontaneous Speech (IPDS, CD-ROMS 1995, 1996, 1997), the mean duration of silent intervals at the wb in mergers was only little longer. The corpus showed mean stop closure durations between 38 and 64 ms, see Figure 2.1. The competitive situation between intended silent intervals and those caused by stop closures makes it impossible to distinguish them and puts the usability of silent intervals as wb markers into question - at least in the durational range we have dealt with so far.

3.4.4. Comparative vowel duration

In section 3.3.4 we put forward the hypothesis that phonemically identical sequences coming from syntactically different constructions might be distinguishable through opposing vowel durations, e.g. short-long for noun-noun compound Augenmaß/sense of proportion vs. long-short for noun-verb phrase [mit den] Augen-maß/measured with the eye or vice versa. Rietveld (1980) had observed that French speech showed such a pattern. His findings, however, were not transferable to our data of German speech. Phonemically identical word pairs were analysed regarding the duration of their primary and secondary stressed vowels or diphthongs. We analysed 160 items produced by different speakers, 80 natural words and their merged counterparts. In the above mentioned example Augenmaß - Augen\_maß for instance, the duration of diphthong /au/ and vowel /a/ were measured. All natural items had the primary stressed vowel in the first syllable, the secondary stressed vowel was in the ultimate or penultimate syllable position. The ratio of the duration between primary and secondary stressed vowel was calculated (vowel 1/vowel 2). Linear mixed-effect modelling was applied with vowel ratio as the dependent variable, boundary condition (natural/merged) as independent variable and participants and items as crossed random factors; outliers were removed. No effect for boundary condition was found (p > 0.6).

3.4.5. Contact geminates - test 1

The reading texts described in section 3.2.2 also formed the source from which the geminate-singleton targets were taken. The targets had not been placed deliberately in the texts. Rather, we became aware of the potential of contact geminates when labelling the acoustic data. Fifteen target words were analysed across speakers.
3. Production experiment

Targets were either a single word to derive initial, medial or final nasal singletons /n/ and /m/ from like *mager/skinny, Monat/month, Guthaben/credit balance* (measured segments are shown as bold) or a phrase of two words that contained nasal contact geminates /nn/ or /mm/ as in *schon_nach/already after*. Sometimes, a target contained both nasal singletons and contact geminates: *einem_Monat/one month, dat*. As a rule, all nasal durations within a target were measured (see Appendix, section A.1).

Furthermore, we included /nm/ as a quasi-contact geminate as in *einem_Mann/a man, acc.*. In contrast to a homorganic contact geminate, which consists of two succeeding homorganic phonemes, the term quasi-geminate shall describe that two akin phonemes meet, which share the manner of articulation (nasals), but their place of articulation is not identical. In the remainder of this thesis, the term “contact geminate” covers both types of contact geminates.

Since lexical frequency might influence the timing characteristics of segments (Pluymakers, Ernestus, and Baayen, 2005), this factor was considered during the analysis. Lexical frequency was determined according to the frequency data provided by Wortschatz Universität Leipzig (http://wortschatz.uni-leipzig.de/) which comprises 35 million sentences taken form publicly accessible sources (mostly newspapers). For each word, the corpus gives two results: total frequency of the lemma and frequency class which is calculated through comparison with the most frequent German word (masc. article *der*) as reference. The calculation of frequency classes out of total frequency is done logarithmically and follows Zipf’s law. The law states “that the rank of a word (in terms of its frequency) is approximately inversely proportional to its actual frequency, and so produces a hyperbolic distribution” (Tullo and Hurford, 2003). “Der” represents frequency class 1. Words up to frequency class 16 are regarded as common words while words belonging to frequency class 17 and above are considered as being outside the general linguistic usage (Quasthoff, 2009, p. 30). Frequency classes are dynamic, because new words enter a language while known words might be less often used due to various reasons (e.g. fashionable terms disappear). The highest frequency class of the corpus is 21 with extremely rare words.

---

14 Sometimes, quasi contact geminates were assimilated. The cited phrase *einem_Mann/a man, dat.*, for instance, was occasionally uttered as *eim_Mann*. Quasi-geminates which were assimilated to homorganic geminates were counted as homorganic geminates.

15 A word that belongs to frequency class 21, for instance, is the rarely used noun *Schwippschwager* which means *brother of one’s brother/sister in law or husband of one’s sister in law*. The word *Schwippschwager* has only one count in the corpus of Wortschatz Universität Leipzig. The reference word *der* representing frequency class 1 appears $2^{21}$ times more often than *Schwippschwager* in the corpus.
3. Production experiment

Our target words reached from frequency classes 3 to 14 and thus belong to the area of commonly used words. Since two words are necessary to form a contact geminate, the frequency of the first and the second word were averaged. In sum, 115 nasal contact geminates (12 /nn/, 33 /mm/, 70 /nm/) and 166 nasal singletons (35 initial, 68 medial and 63 final) were suitable for analysis (total: 281 items). Thirty-three of the final nasals were followed by an [IP] boundary, which might condition lengthening of the segment. In our statistical analysis, duration was the dependent variable, condition (initial singleton, medial singleton, non [IP] final singleton, [IP] final singleton; geminate) and lexical frequency were the independent variables and participants and items were the crossed random factors. Outliers were removed. The results showed an effect of initial strengthening, as well as word-final lengthening. Singletons in the word-medial condition were shorter than singletons in word-initial and -final conditions (medial-initial: beta = 0.011, lower bound = 0.0003, upper bound = 0.027, p = 0.052; medial-final: beta = 0.028, lower bound = 0.0025, upper bound = 0.032, p = 0.072). Both effects approached significance. For word-final singletons at [IP] boundaries, the effect of word-final lengthening compared to medial singletons was highly significant (beta = 0.056, lower bound = 0.049, upper bound = 0.075, p < 0.0001). There was no effect of lexical frequency.

[Figure 3.4. Condition-induced nasal singleton durations compared to all pooled nasal contact geminates.]

Initial and non [IP] final singletons showed no significant difference in duration (beta = 0.017, lower bound = -0.017, upper bound = 0.023, p > 0.7).

16a Beta” denotes the estimate adjustment to the intercept, “lower bound” and “upper bound” denote the lower and upper bounds of the highest posterior density interval for 95 per cent of the probability density.
3. Production experiment

[IP]-final singletons were significantly longer than non-[IP]-final singletons (beta = 0.032, lower bound = 0.067, upper bound = 0.027, p < 0.0001). The duration of word-final singletons at [IP] boundaries did not differ from those of geminates (beta = 0.002, lower bound = -0.016, upper bound = 0.007, p > 0.4).

Except for [IP]-final singletons, all singleton conditions differed in duration to geminates (initial: beta = 0.047, lower bound = 0.057, upper bound = 0.031, p < 0.0001, medial: beta = 0.057, lower bound = 0.066, upper bound = 0.048, p < 0.0001, non-[IP]-final: beta = 0.029, lower bound = 0.059, upper bound = 0.025, p < 0.0001). Durations of singletons and geminates are visualised in Figure 3.4. For lexical frequency, neither an interaction with condition nor a main effect was found.

For all speakers, the duration of the longest singleton (mostly [IP]-final ones) exceeded the shortest contact geminate. In order to check the potential of contact geminates as wb markers, we looked for the point at which our nasals were best dividable into singletons and geminates. We assumed an automatic classification, taking only the duration of nasals as a criterion, and found that 77 per cent of the items were correctly identifiable as singletons and contact geminates, if the separation boundary was fixed at 100 ms. This observation shows that singletons and geminates share some overlapping area of duration. An ideal classification of all items is hardly possible despite the fact that the singleton and geminate group have clearly differing mean durations (mean singletons: 87 ms, mean contact geminates: 131 ms). However, if the separation boundary is carefully chosen, the error rate can be minimised. There is yet another method to improve the classification results, namely through the identification of singletons at [IP] boundaries. Their durations are usually similar to those of contact geminates. In an automatic classification process that sorts items according to segment duration, lengthened [IP]-final singletons would end up in the geminate group. But there is a way to identify them. The breath group theory of Lieberman (1967) assumes that speakers tend to take a breath after the end of a syntactic unit and thus insert a silent interval. Hence, a long nasal (for our dataset, “long” means >= 100 ms) followed by a silent interval had to be a lengthened singleton. Due to phonotactic rules of German speech, a silent interval cannot follow after a nasal contact geminate. The second time slot of the contact geminate equals the first phoneme of a word. In the case of German nasals, only vowels can follow at the word onset (Yu, 1992). From an articulatory point of view, the vocal folds vibrate uninterruptedly when the sequence nasal + vowel is produced (einem Mann: nasal + vowel /m:a/). Consequently, a check of the right neighbourhood of all nasals >= 100

75
ms was carried out in order to find those items that were followed by a silent interval. The procedure further improved the result for correct identification of nasal items in our dataset by 9 per cent - these are the newly covered IP-final singletons that were followed by a silent interval. The total identification score for all items in the automatic classification could thus be raised to 86 per cent.

The question of segment duration according to condition leads us to the singleton-geminate ratio, which was 1.51. If the quasi-geminates (mean: 143 ms) were taken out and only homorganic geminates (mean: 112 ms) were included in the calculation, the ratio decreased to 1.29. Quasi-geminates and homorganic geminates differed significantly in their durations (beta = 0.028, lower bound = 0.014, upper bound = 0.046, p = 0.002), see Figure 3.5.

What conclusions can be drawn from the obtained results? Most importantly, nasal contact geminates proved to be useful cues to word boundaries in German speech. When two nasals framed a \textit{w HB} which means that one word ended and the next one began with the same nasal, the phonemes were usually acoustically concatenated and uttered as one lengthened segment. This finding is in accord with Lea (1980) who analysed English sibilants. He explained that the exceeding of a certain threshold in the sibilant’s duration might signal a division point of two succeeding words. We found that contact geminates were significantly longer than initial, medial, and final singletons without an
3. Production experiment

Furthermore, we obtained segment-initial and segment-final strengthening effects. Lengthened word-final singletons at [IP] boundaries reached an average duration similar to contact geminates. They are identifiable when followed by a silent interval.

The quasi-geminate /nm/ was included in the study and treated like the regular ones. It was sometimes assimilated, but not as a rule. This has to be attributed to the fact that data elicitation was based on reading, which is usually characterised by a more careful articulation compared to spontaneous speech.

Our singleton-geminate ratio was below the values cited in the literature for languages with consonantal length contrast (1.61 to 2.61). While durations for /nn/ and /mm/ proposed genuine consonant lengthening resulting in a geminate, durations for /nm/ rather suggested the concatenation of two singletons. Astonishingly, our ratio was also below the ratio reported by Mikuteit (2007). She had found a geminate-singleton ratio of 1.82 for German alveolar stop geminates. One explanation for this divergence might be found in the material (prefixed verbs, e.g. mittesten/to test along with) and reading style of the test persons. Mikuteit’s participants had been instructed to utter isolated words with a monotone intonation, our targets were embedded in sentences. Sentence prosody might have influenced segment duration. The diverging ratios might also be due to the fact that different phoneme characteristics were observed: closure duration for stop geminates and sonority in the case of nasals.

In order to investigate the timing characteristics of singletons and contact geminates further, a second production experiment was carried out in which the number and condition of each singleton segment was controlled (initial, medial, non [IP]-final, [IP]-final), as well as the number of homorganic and quasi-geminates.

### 3.4.6. Contact geminates - test 2

In order to replicate test 1 under more controlled conditions, an additional recording became necessary. The methods are described in the following; results and discussion conclude the section.

---

17 Möbius and van Santen (1996) give segment durations of nasals according to their position within the syllable. Onset: /n/ = 52 ms, /m/ = 57 ms; ambisyllabic: /n/ = 56 ms, /m/ = 71; coda: /n/ = 77 ms, /m/ = 85 ms, source: The Kiel Corpus of Read Speech, as published on CD-ROM, IPDS 1994. The mean duration of our quasi-geminate was 143 ms.
3. Production experiment

Methods

Participants

Eleven speakers took part in the recordings. One of the speakers had also taken part in the previous recording for the production experiment, the others had not. Test persons were between 22 and 52 years old; 8 of them were female, 3 male. Six persons spoke dialect (Saxon, Swabian, Baden dialects), the others spoke SG. The group comprised of academics and non-academics. The recording took place at two locations - Rostock and Konstanz. Six participants had no linguistic background, five were students of linguistics.

Materials

Forty-eight sentences were constructed in the form: noun-verb-noun, e.g. Mechthild spielte Geige/Mechthild played the violin or noun-verb-abverb e.g. Wilhelm mogelt öfter/Wilhelm cheats quite often; the initial noun was always a first name. Singletons /n/ and /m/ were placed in the four conditions of interest: word-initially, word-medially, word-finally but no IP boundary, word-finally at an IP boundary. For each condition and phoneme, 4 items were prepared. This gives 32 sentences for the singleton items (2 phonemes x 4 conditions x 4 items). As far as possible, they were also put in different vowel contexts. Word-initial /n/, for instance, appeared in the vowel surroundings /no/, /na/, /ni/, and /ne/ (Norbert, Nadja, Nicki, Nele). For each sentence, only one nasal was analysed. Word-initial and -medial nasal singletons were taken from the first word of the sentence (e.g. Mechthild, Linda). Non IP-final singletons without an IP boundary were also taken from the first sentence constituent (e.g. Wolfram). For word-final singletons at an IP boundary, the last word of the sentence was used, which was also a first name (e.g. Gerhard hörte Adam/Gerhard heard Adam). In some sentences, adverbs were in sentence-final position - like in the above mentioned example Wilhelm mogelt öfter/Wilhelm cheats quite often. However, adverbs were not used to derive nasal segments from. Contact geminates were formed through the concatenation of the last segment of the first word and the first segment of the second word (e.g. Wilhelm mogelt öfter/Wilhelm cheats quite often). In this experiment, quasi-geminate /mn/ (Adam nuschelt ständig/Adam mumbles all the time), which had not been tested previously, was included. For the geminate condition, 16 token were placed, four for each contact.
3. Production experiment

geminate. Reading lists were prepared which contained all 48 sentences in a pseudo-randomised order. All speakers were recorded twice, which means that in total, 96 items could potentially be produced by each participant (32 singleton items + 16 geminate items times two recordings).

As in test 1, lexical frequency was coded in frequency classes according to Wortschatz Universität Leipzig. For details about the procedure, see section 3.4.5. Lexical frequency of the targets was not controlled. The targets ranged between frequency classes 8 and 21. Three rarely used first names - Oswin, Irntraut, Nicki - belonged to the group of uncommon words (frequency classes 21, 19, 17), all other targets had frequency classes that refer to the general linguistic usage.

Procedure

Recordings were made at two locations: (1) in an anechoic chamber at the University of Music and Theatre Rostock’s recording studio using a Audio Technica AT4050 SM microphone and an Foster FR-2 Digital Recorder (48 KHz/24 Bit); (2) in a soundproof room at Konstanz University using a MXL 990 condenser microphone and a TASCAM HD-P2 Portable High-Definition Stereo Audio Recorder (44.1 KHz/24 Bit). Test persons were instructed to read the sentences aloud with a consistent declarative intonation. Each speaker had one individual recording session in which he or she read the material twice.

Results and Discussion

Target segments from all recordings were cut and labelled. Singleton /n/ and /m/ and contact geminates /nn/, /mm/, /mn/ and /nm/ were measured. Items were coded (1) according to condition: initial singleton, medial singleton, non-final singleton, final singleton, contact geminate and (2) according to type: /n/, /m/, /mm/, /nn/, /nm/, /mn/.

Speaker KB from Hamburg, who had produced the sentences in a very slow speech tempo, was excluded from analysis. The mean duration for geminates over all remaining speakers was 127 ms, standard deviation 21 ms. KB’s mean geminate duration was 184 ms, which was 2.5 times above standard deviation.
Apart from KB’s production data, a few other items had to be excluded as well. These were either geminate targets in which the neighbouring nasals had not been produced as a contact geminate but as two nasals singletons or singleton targets in which the actual singleton was missing, e.g. Marti- instead of Martin, or mispronounced, e.g. Irmgard instead of Irngard. Finally, 301 nasal contact geminates (73 /nn/, 73/mm/, 77 /mn/ and 78 /nm/) and 630 nasal singletons (/m/: 80 initial, 76 medial, 78 non IP-final, 78 IP-final; /n/: 79 initial, 80 medial, 80 non IP-final, 79 IP-final) remained for analysis (total: 931 items).

Results were analysed using linear mixed-effect modelling. We analysed segment duration with segment condition and segment type, recording (first/second) and lexical frequency as fixed factors, segment duration as the dependent variable and participants and items as crossed random factors. Outliers were removed.

There was an effect of recording. Items from the second recording were of significantly shorter duration than segments from the first recording (beta = 0.006, lower bound = 0.008, upper bound = 0.003, p = 0.0002).

No interaction and no main effect for lexical frequency were found.

Geminates (/mm/, /nn/, /nm/ and /mn/) showed no differences in duration. In contrast to the shortest geminate (/nn/, mean: 113 ms), we found no effect for segment type (all p-values > 0.4). Other than in the first experiment, quasi-geminates did not differ from homorganic geminates in duration, see Figure 3.6. This might be due to the

![Figure 3.6](image)

**Figure 3.6.** Duration of homorganic and quasi-geminates in two tests. The first test (top) comprised only one quasi-geminate (number of tested items: 12 /nn/, 33 /mm/, 70 /nm/). The second test included both quasi-geminates and more items in general (73 /nn/, 73/mm/, 77 /mn/ and 78 /nm/).
facts that (1) a greater number of items were analysed and (2) other factors that might have influenced segment duration in the first test (e.g. word/sentence stress) had been controlled.

The usability of contact geminates as markers was verified in the second experiment (see Figure 3.7). Singletons in the initial, medial and non IP-final boundary conditions were significantly shorter than geminates (initial: beta = 0.065, lower bound = 0.071, upper bound = 0.058, p < 0.0001; medial: beta = 0.032, lower bound = 0.039, upper bound = 0.025, p < 0.0001; non IP-final: beta = 0.039, lower bound = 0.045, upper bound = 0.033, p < 0.0001). So far, the results of the first test could be replicated. A difference was found in the durational dimensions of IP-final singletons compared to geminates. In the second test, IP-final singletons were significantly longer than geminates (mean durations: 131 ms and 121 ms; beta: 0.012, lower bound = 0.006, upper bound = 0.018, p = 0.0002).

![Figure 3.7. Test 2: Condition-induced nasal singleton duration compared to all pooled nasal contact geminates.](image)

In the first test, IP-final singletons and geminates had similar durations. One reason for the different result in the second test might be that the targets had been produced differently. In the first test, IP-final singletons were partly taken from longer texts (short story, newspaper article) and the IP boundary was not always the sentence boundary. In the second test, speakers read very short sentences in which IP-final singletons had a fixed position - they were always the last segment of the word which concluded a sentence. This might have lead to a stronger lengthening of the last segments.

Singletons also produced results that were in part different to the results obtained in the first test. Generally, singleton types /n/ and /m/ were significantly shorter than the shortest geminate /nn/(/m/: beta = 0.029, lower bound = 0.045, upper bound = 0.012,
p = 0.0004; /n/: beta = 0.025, lower bound = 0.043, upper bound = 0.010, p = 0.002). We obtained rather short initial nasal segments (mean: 56 ms), as well as long medial ones (mean: 83 ms). Consequently, there was no effect of word-initial strengthening (see Figure 3.7). The measured durations for /m/ and /n/ in their respective positions within the word approached those reported by Möbius and van Santen (1996). Initial singletons were shorter compared to singletons in all other conditions; the effect was significant (all p-values < 0.0001). IP-final singletons without an IP boundary were also shorter than medial ones, the effect approached significance (beta = 0.007, lower bound = -0.015, upper bound = 0.001, p > 0.08).

![Item examples per segment condition](image)

**Figure 3.8.** Each segment condition represented by one example item: initial singleton (*Moritz*), medial singleton (*Irmgard*), non-IP-final singleton (*Tristan*), IP-final singleton (*Heidi kannte Wilhelm*); geminate (*Martinickte*). The longest segment durations and greatest variation were to be found in IP-final singletons.

Final singletons at IP boundaries had the longest durations of all analysed items, as mentioned above. In comparison to medial singletons, this produced a significant effect of word-final lengthening (beta = 0.043, lower bound = 0.035, upper bound = 0.052, p < 0.0001). Figure 3.8 illustrates the temporal organisation of the analysed items according to condition. Obviously, the reason why the singleton group had produced different effects in the second test compared to the first one was a differing temporal organisation of medial singletons. In the first analysis, they had been shorter than initial singletons and final ones without an IP boundary.

Let us have a closer look at these long medial singletons and the words they have been taken from. Medial singletons always originated from the first syllable of the sentence-
initial word where they were positioned in the coda of a stressed syllable (e.g. *Linda*). This circumstance obviously produced an effect of syllable-final lengthening. To investigate the matter further, targets with medial nasal singletons in the syllable-final condition (*n* = 156) were contrasted to targets with medial singletons in the syllable-initial condition, such as *Hermann* (*n* = 116). All speech material stemmed from the same recordings. The segments containing medial singletons in the syllable-initial condition originated from words which had already been used for the analysis of another nasal segment. As mentioned before, each target word was designated for the analysis of one segment. *Hermann*, for instance, had been used for the analysis of a contact geminate (*Hermann nähte/Hermann sewed*). Since the target coincidentally also contained medial singleton /m/ in syllable-initial position, it was re-used for the test of medial singletons in different conditions. For a list of all used segments, see Appendix A.2.2.2.

We analysed segment duration using linear-mixed-effect modelling with segment duration as dependent variable, segment condition (syllable-final/initial), segment type (/n/, /m/) and recording as fixed factors and participants and items as crossed random factors. Outliers were removed. There was neither an effect of recording (*p* > 0.8), nor of segment type (*p* > 0.3). There was no interaction of the fixed factors (all *p*-values > 0.2). We found an effect of condition: syllable-initial medial singletons were significantly shorter than medial singletons in the syllable-final condition (beta = 0.0232, lower bound = 0.037, upper bound = 0.0098, *p* = 0.0024), see Figure 3.9.

![Figure 3.9. Durations of word-medial /m/ and /n/ according to condition.](image)

Compared to the first test, the geminate-singleton-ratio decreased (1.36 in test 2, 1.51 in test 1). This effect was influenced by (1) lengthened medial singletons in the syllable-final condition and (2) IP-final singletons that were longer than geminates.

---

18 The newly-labelled medial singletons in the syllable-initial condition were not included in the calculation of the ratio, only the original items.
3. Production experiment

3.5. Production experiment - summary of the results

The production experiment had the aim to analyse the characteristics of unpredictable \textit{wb} markers (glottal stops/creaky voice, stop allophones, silent intervals, comparative vowel duration, contact geminates). Word pairs were designed that consisted of a natural word and a phonemically identical merged twin that included a \textit{wb}, e.g. \textit{Monat/month} and \textit{Mona_T["olle]}/first and family name. The merger was formed through the embedding of words or word parts (cf. Lehiste, 1965; Rietveld, 1980; McQueen, Norris, and Cutler, 1994; Christophe et al., 2003). Nineteen speakers, among them nine with dialects from five dialect regions of Germany, were recorded reading texts in which natural words and their phonemically identical counterparts were distributed. Targets were cut and labelled with Praat.

Glottalisation of word-initial vowels occurred frequently: two-thirds of the targets were affected. Creaky voice was the most common realisation. The finding reflects the results of Wesener (1999). A sole glottal stop as \textit{wb} marker was less often produced, the combination glottal stop + creaky voice was rare. Wesener had reported that a glottal stop usually appeared in unison with creaky voice. This result could not be replicated with our dataset. We can only speculate about the reasons for this discrepancy. One might be that \textbf{SG} speakers produce the glottal stop (and thus glottal stop-creaky voice combinations) more often than dialect speakers, because \textbf{SG} has more boundary signals in general. However, the number of the analysed items within our dataset was too small to come to a reliable conclusion, further investigations are needed.

Contact geminates proved to be reliable \textit{wb} markers. Exemplarily, nasal segments were tested. In German, where geminates are not part of the phonemic inventory, a segment might be geminated when one word ends and the following one begins with the same phoneme. In two tests, nasal contact geminates were found to be significantly longer than nasal singletons in the initial, medial, and word-final but not \textbf{IP}-final condition. The first test resulted in a nasal singleton-geminate ratio of 1.51 which decreased in the second test to 1.36 caused by an increased number of lengthened medial singletons as well as \textbf{IP}-final singletons that were longer than geminates. Long medial singletons in the second test arose through a syllable final position which conditioned segment lengthening. We also proposed a method for an automatic classification of items into singletons and geminates using segment duration as well as silent intervals in the right neighbourhood as criteria. Silent intervals proved to be very useful to identify lengthened...
Stop allophones were examined regarding segment intensity. Intensity values did not differ, no matter at which position of the word (initially, medially, finally) the stop was located.

Silent intervals were measured in all target words of the corpus in which they appeared except for the singleton-geminate items. Thereby, only the wb area in the merged word and the corresponding location in the natural counterpart were considered. Even though silent intervals in mergers had a slightly longer average duration (76 ms) than those in natural words (60 ms), the effect was not significant. In merged targets that contained a stop after the wb (respectively, the corresponding area in natural words), silent interval durations ranged in the dimension of normal stop closure times. It was impossible to say if the measured silence was deliberately inserted or governed by articulatory processes involving stop articulation (cf. de Pijper and Sanderman, 1994).

The phenomenon of comparative vowel duration which Rietveld (1980) found for French, could not be replicated with our German targets. Rietveld had reported that French speech material comprising phonemically identical but syntactically different sequences had opposing vowel durations. A noun phrase le muscat perdu/the spoiled muscat (wine) could be distinguished from a phonemically identical sentence le musc a perdu/the musk is evaporated through the temporal organisation of the full vowels. They showed opposing duration patterns (long-short, short-long) in the respective sequences. We examined whether our speakers also used opposing vowel durations to disambiguate phonemically identical German word pairs such as Augenmaß/sense of proportion - [mit den] Augen_mäß/measured with the eye, but they did not.

The analysis of our production data allowed first insights into the characteristics and usability of wb markers. In the following chapter, we pursue the question which markers are important for the perceptual identification of a wb.
4. Perception experiment

4.1. Introduction

The aim of the perception experiment was to examine whether \textit{wb} markers are perceptible. Listeners were asked to judge if audio samples sounded natural, meaning, like a normal word. We assumed that this naturalness judgement depended on two conditions, namely (1) on the number of \textit{wb} markers present and (2) on the individual boundary signalling strength of each marker.\footnote{The criterion “number of \textit{wb} markers” relates to the examined ones. We cannot exclude that there were other but unaccounted markers present in our targets.} In order to test these conditions, a perception experiment was conducted that consisted of two parts. The first part focussed on \textit{wb} marker quantity, the second one on quality. In both parts, test persons heard audio samples - a set of natural and merged items, but never the natural and the merged version of a word in one set.

Note that the material used in perception experiment 1 stems from the recordings made for the production experiment, where various speakers had read texts containing word pairs. Out of this corpus, stimuli produced by four speakers were chosen. For perception experiment 2, new targets were generated. Only one speaker provided the speech data; he had not taken part in the recordings for the production experiment. One special characteristic of the perception experiment is that the majority of our participants had no linguistic background. Only a few listeners had undertaken speech training as part of their professional education. Thereby we were able to check if the performance of the speech-trained participants differed in comparison to the linguistically naive listeners. A generalisation, however, was not possible because the two groups were too unbalanced in numbers.

Similar to the production experiment, participants of the perception experiment came from various dialectal regions of Germany as well as from \textit{SG} areas. We therefore used
this opportunity to check if the performance of listeners was influenced by affiliation to a certain dialect group. This analysis was done in perception experiment 1. Since we could not find an effect, the issue was not pursued further in experiment 2.

4.2. Experiment 1

The first part of the perception experiment (in the remainder: experiment 1) aimed to give first answers to the question if the presence of several [wb] markers in a target improved the identification of a [wb] in comparison to targets with only one. The audio material consisted of sequences that were cut out of the audio stream. These sequences were not manipulated. Our aim was to work with completely natural language and to preserve all markers as they had been produced by our speakers. Targets carried one or a combination of two or three of the following [wb] markers:

1. glottal stops
2. stop allophones
3. /l/-allophones
4. silent intervals
5. non-standard stress patterns

Since we were dealing with unmanipulated items, the following features were also regarded as [wb] markers:

6. vowel quality/quantity
7. prosodic cue

Most of the [wb] markers analysed in the production experiment were now tested with regard to their perception. Creaky voice, however could not be examined further. The audio material of the four speakers chosen for the perception experiment contained only one sample (b_All) that had the combination glottal stop + creaky voice. All other target words contained the glottal stop as a single marker.

---

2There were also too few participants per dialect group to answer the question satisfactorily. However, we were able to get some first ideas about the influence of dialect on perception tasks. For a comprehensive study, a much larger group of listeners had been necessary which was beyond the scope of the experiment.
Contact geminates were also not tested in the perception experiment. The production experiment has shown a durational contrast between contact geminates and singletons in word-initial, medial and final condition except when the latter one was also the IP-final condition. Contact geminates and IP-final singletons showed no durational contrast; phrase-final lengthening of the final singleton has produced similar and even longer durations compared to contact geminates. Since both conditions are natural in German, we did not expect an “unnatural” judgement for contact geminates in the perception task.

The third marker that was not tested in this first perception experiment part was comparative vowel duration, because the production experiment had demonstrated that there is no such marker in German. Targets, however, had to contain at least one wb marker.

Similar to the comparative vowel duration samples, items with stop allophones had neither produced an effect in the production part, but they were kept for the perception experiment for three reasons. First, the results obtained in the production experiment showed a high standard deviation. Second, aspirational cues had not been tested directly because raw values were found to be difficult for comparison, see section 3.2.3. Third, stimuli had been created in such a way that mergers with allophones in wrong word positions showed well audible differences in contrast to the natural counterparts. For instance, the target Koda_k/name of a company with a highly aspirated, fully released final stop gave the impression of unnatural hyperarticulation.

Silent intervals were also tested in the perception experiment despite the fact that the production experiment had generated no significant effect. The mean silent interval duration at the wb in mergers had been 76 ms. In the perception experiment of de Pijper and Sanderman (1994) listeners evaluated a silent interval as an audible boundary if it was longer than 100 ms (section 2.8). Expecting to replicate this result, targets with silent intervals around 100 ms were included in our experiment.

Two new markers were included in the perception experiment: /l/-allophones (Pappel/-paspel - Pappe_l[liegt]/cardboard lies) and non-standard stress ([dass Wind und] ,Wasser `wirtschaft[l ich sind]/that wind and water are profitable, standard stress: `Wasser,wirtschaft/ water economics).

3The sequence was cut out of the sentence Besonders die Koda klingt toll/Especially the coda sounds great.
4. Perception experiment

Vowel quality/quantity, e.g. of /a/ in Mona\textsubscript{T} - Monat (originating from an open vs. closed syllable) had not been controlled in this first experimental part, but was considered as an additional \textsubscript{wb} marker. Furthermore, seven word pairs had a prosody cue with respect to word position within the utterance. This means that the natural word had been realised with an intonation contour different from the one of the merged word, e.g. Monat in phrase-final position with a falling nuclear tone and Mona\textsubscript{T} in phrase-initial position with rising nuclear tone. Like vowel quality and quantity, a prosody cue was regarded as a \textsubscript{wb} marker.

Within the frame of the experiment, dialectal parameters were observed as well. The presence of dialect speakers in the group of listeners and one in the group of speakers who provided the acoustic stimuli provided an opportunity to check upon dialect-specific tolerances. Our aim was to get a first impression, an in-depth study would have required more participants. Data from Saxon, Baden and Swabian listeners were used for this task. Their respective dialects show no distinction between [+ fortis] and [- fortis] stops (Kohler, 1995): /p, t, k/ are articulated as /b, d, g/, e.g. klein/little $\rightarrow$ glein. Through assimilation processes across word boundaries some of the presented targets had undergone lenition of the word-initial [+ fortis] stop, e.g. target k\textsubscript{lebt} coming from the sequence [das Handwerk]k\textsubscript{lebt}/craftsmanship is alive was produced as [glebt], the natural word is klebt/sticks, 3. p. sg. The question was if these dialectically coloured samples would be tolerated as sounding “natural” by the aforementioned listener group.

We also wanted to learn if our listeners with Swabian or Baden dialects who are accustomed to stress patterns which differ from SG (e.g. ‘Büro/office, SG: B\text{"u}ro; Tun\text{"e}l/tunnel, SG: ‘Tunnel), perceive targets with non-standard stress as natural.

4.2.1. Methods

4.2.1.1. Participants

Fifty test persons took part in the experiment, 27 female and 23 male. They were aged between 19 and 62 and came from Bavaria, Baden-Württemberg, North Rhine-Westphalia, Rhineland-Palatinate, Saxony, Thuringia, Saxony-Anhalt, Berlin, Branden-

\footnote{Consonantal quantity of fricatives in merged fricative-stop clusters was not considered as an additional \textsubscript{wb} marker, because durations hardly differed from the segments in the natural counterparts (max. difference: 15 ms).}

\footnote{The seven word pairs in question are listed in the Appendix, section A.2.1.1.}
4. Perception experiment

burg, Lower Saxony, Hamburg, Schleswig-Holstein and Mecklenburg-West Pomerania. Our listener pool thus covered nearly all German States (13 of 16). Sixteen listeners were dialect speakers (Bavarian, Swabian, Baden, Saxon, Cologne, Oberlausitz, Palatine dialects); all others spoke SG. Listeners were academics and non-academics, but none of them had a linguistic background. Speakers of the stimuli and listeners were not acquainted with each other.

4.2.1.2. Materials

Targets were taken from the speech corpus generated for the production experiment. Stimuli uttered by four participants of the production experiment were selected. These were one male and two female speakers of SG as well as one male dialect speaker (Saxon). The four SG speakers had produced most targets with a well-audible difference between the natural and the merged word, which especially applied to the items with stop allophones, see also 3.4.2.

The experiment comprised 68 targets - 34 natural words and their merged counterparts (Monat - Mona_T). Remember that all items were unmanipulated. Natural words as well as the respective phonemically identical sequences had simply been cut out of the speech stream. Mergers had one, two or three wb markers. For more detailed information, also regarding the incorporation of wb markers, see section 3.2.2. Lexical frequency of the targets was not controlled. Two sample sets with acoustic stimuli had been created. Each set contained 34 targets: natural and merged words in a pseudo-randomised order which was done manually. Neither set contained the natural and the merged version of a word. Natural word Monat, for instance, belonged to set 1, merger Mona_T to set 2. All other word pairs were divided accordingly between the two sets. The acoustic stimuli that were played to the participants had been arranged in such a way that each of the four speakers were heard en-bloc. The SG speakers came first, the dialect speaker last.

Through assimilation effects across the wb, the merging procedure had produced some targets with word-initial [- fortis] stops where SG demands a [+ fortis] one. If /k/, for instance, was the last phoneme of a word and the following one started with an /l/, the stop was assimilated to the lateral across the wb (cf. Kohler, 1995; Wiese, 2000). As a result, the stop lost its feature [+ fortis], e.g. [das Handwerk]k_lebt/craftsmanship is alive changed to [glebt], dialectal form of klebt/sticks, 3. p. sg. Besides those artificially arisen
4. Perception experiment

items, each set contained naturally produced dialect targets, provided by the speaker from Saxony.

Items with “wrong” stress patterns arose in two ways. 
(1) Two unstressed syllables were merged to a new word that had no stress at all, e.g. an article and a prefix as in die_Be (original context: die Belüftung/the ventilation). The natural twin 'Diebe/thieves' has stress on the first syllable. Merged targets which had been formed according to this scheme always carried at least two \( \text{wd} \) markers, here: stress manipulation + word-initial stop allophone /b/ in word-medial position. 
(2) Two words, each with primary stress, were merged to a new word. One of the words also carried the sentence stress. Perceivably, this resulted in an impression of “stronger” stress compared to the other primary-stressed word. The merger Wasser_wirtschaft/water economics, for example, was taken out of the sentence Dass Wind und Wasser wirtschaftlich sind, wird von Befürwortern der Kernenergie verneint/The fact that wind and water are profitable, has been negated by advocates of nuclear energy. Both 'Wasser/water' and 'wirtschaftlich/profitable' carry primary stress. The first syllable of 'wirtschaftlich', however, also carries the sentence stress of the relative clause (represented through the bold font). If 'Wasser' and 'wirtschaftlich' are merged to 'Wasser_wirtschaft', stress on syllable /wirt/ is perceived as stronger than stress on syllable /was/. The result is the auditory impression of stress shift.

4.2.1.3. Procedure

The experiment was carried out with an Acer Travelmate 525 TX laptop and Sony MDR CD 550 headphones which were plugged into the laptop. The volume was adjusted at a comfortable level. Test persons sat in front of a laptop and wore headphones which were plugged into the computer. The laptop screen showed the soundplayer which had to be operated with a computer mouse. At the beginning of the experiment, the player showed only the START button. After clicking the start button, the first audio sample was played. From now on the soundplayer displayed two buttons: ERNEUT HÖREN/Repeat and WEITER/continue. Audio samples could be replayed twice. After two replays, the button was deactivated and the listeners could only go on by pressing

---

6 According to Selkirk (2008) the two unstressed syllables in die_Be have a different status, one is stronger than the other.
7 We thank Silvio Robel.
the WEITER/Continue button. The participants were instructed to listen to the audio sample and to decide if the target stemmed from sentence 1 or 2, which were given in an A5-booklet that lay in front of them. The booklet showed the natural and the merged word embedded in their original sentences. For each word pair, one page was used. Both targets were underlined, as the following example illustrates (translations are only given here, they were not in the booklet):

Der Ball wollte einfach nicht ins Tor gehen/The ball would simply not find its way to the goal.

Der Zug fährt 7.13 Uhr ab Allensbach/The train goes from Allensbach at 7.13 am.

(For the complete list of stimuli, see Appendix section A.2.1)

Listeners should mark the chosen sentence with a cross. In this self-paced experiment, test persons were free to decide whether they first read the sentences in the booklet and listened to the target word afterwards or to do it in the reverse way. Participants were asked to come to a decision, but they should not guess. If they had indeed no idea how to answer, they should leave the respective booklet page blank.

The focus of the experiment was on phonetics. Words were heard out of context to disable or at least weaken syntactic, semantic or morphological recognition strategies. Therefore, test persons were given the two concessions mentioned before: the booklet with target-related context and the possibility to replay each sample twice.

Each test person heard only one set of acoustic stimuli (34 items). All listeners were tested individually in a quiet room at the University of Applied Sciences Konstanz or in quiet private rooms. Before the actual experiment started, a training session with three audio samples was conducted. Training samples comprised two mergers and one natural word and had been produced by a female speaker from Saxony (Dresden). The training session should familiarise listeners with the acoustic nature of the targets and acquaint them with the operation of the sound player. If questions arose during the test phase, the experimenter answered them. Then, the experimental file was loaded into the soundplayer and the participant was free to start.
4. Perception experiment

4.2.2. Hypotheses and research questions

**Single marker targets vs. multi marker targets**

We assumed that the number of **wb** markers present in a merger influenced recognition. Targets with one marker, for instance [hal]b-rauchen (word-final stop allophone in initial position, natural word: brauchen/to need) would be more difficult to detect than those with several, for instance Mona_T[¨olle] (word-initial stop-allophone in final position, vowel quality/quantity, prosody cue; natural word: Monat/month).

**Dialect-specific tolerance**

It was expected that dialect speakers from Saxony, Swabia and Baden would tolerate dialectal colourings which are present in their own respective dialects, in particular the lenition of [+ fortis] stops and non-regular word stress. Mergers like [das Handwer]k_lebt/craftsmanship is alive (natural word: klebt/sticks, 3. p. sg.) which had come up as [glebt] due to an assimilation process across the **wb** would thus be judged as natural words by speakers whose dialects are characterised by regular lenition of [+ fortis] stops. A similar mechanism was assumed for the samples with non-standard stress. Mergers with non-standard stress would be tolerated as natural words by listeners from Baden and Swabia who might have a greater tolerance towards non-regular stress in their dialects. Beside the influence of dialect items on dialect speakers we also examined, which impact dialect samples - delivered by the speaker from Saxony - have on listeners with **SG** or dialects. We hypothesised that the acoustic stimuli delivered by the Saxon speaker were more often judged as mergers, because his items were mostly those with [+ fortis] stop allophones which he had produced as [- fortis].

**Silent intervals**

Assuming that Dutch listeners perceived silent intervals above 100 ms as an audible boundary (de Pijper and Sanderman, 1994), we expected that silent intervals with durations around this value would function as **wb** markers. In other words: Targets with silent intervals of around 100 ms at the **wb** will lead the participants to identify these items as mergers.

Research Questions for the first experimental part were:

1. Does recognition of word boundaries improve with an increasing number of boundary markers?
4. Perception experiment

2. Is it possible to allocate a minimum duration for silent intervals which makes them effective as boundary signals?

3. Is dialect an interfering factor in the process of \textit{wb} recognition?

4.2.3. Results

The analysis is based on 1650 targets, half of them were natural words, half of them mergers. Answers provided in the booklets were manually coded. Lexical frequency was determined according to corpus data provided by Wortschatz Universität Leipzig with frequency class as the corresponding unit (for details, see section 3.4.5). Our targets spanned frequency classes 8 to 19. Remember that words between frequency classes 1 to 16 are regarded as common words. The majority of our targets belonged to this group of common words, only 7 targets (350 items in total) have higher frequency classes and are thus considered to be outside the general linguistic usage (e.g. \textit{Skonto/allowance}, \textit{krauchen/to crouch}).

When none of the sentences had been ticked, the item was counted as incorrect, which happened in 25 cases. The small number reflects the participants’ effort to come to a decision. The undecided cases did not relate to a special \textit{wb} marker group. They comprised both natural and merged items.

The statistical analysis is based on a binominal logistic regression. It allowed us to test the interaction between two and more factors (categorical and continuous). Boundary condition, number of \textit{wb} markers and lexical frequency were the fixed factors, recognition correct (binary: yes/no) the dependent variable, participants and items were the crossed random factors. Natural words were significantly better identified than mergers ($\beta = 0.683$, $z = 5.928$ $p < 0.0001$). Within the merger group ($n = 825$), targets with two or three \textit{wb} markers were more often recognised as mergers than targets with only one (one marker vs. two markers: $\beta = -2.234$, $z = -1.878$, $p = 0.06$; one marker vs. three markers: $\beta = -4.0005$, $z = -2.716$, $p = 0.007$). However, the presence of a third marker did not improve recognition (two vs. three markers: $p = 0.12$). We also found an interaction of number of markers with lexical frequency (two-marker targets: $\beta = 0.169$, $z = 2.178$, $p = 0.0294$; three-marker-targets: $\beta = 0.297$, $z = 2.858$ $p = 0.0043$). This interaction was not investigated further since lexical frequency of the targets had not been controlled and we had unbalanced data, e.g. few low-frequency words.
One \textit{wb} marker that was examined more closely was the silent interval. All targets with silent intervals were multi-marker targets (n=116). This means that apart from the silent intervals, at least one other marker was present at the \textit{wb}. Figure 4.1 demonstrates that silent intervals had probably little or no effect on merger identification. Items that were correctly identified by the test persons had relatively short silent periods. They most likely profited from the additional presence of another \textit{wb} marker like a glottal stop. Even the merger with the longest silent interval of 113 ms (\textit{Koda}[$\text{lingt}$]/\textit{coda sounds}, natural word: \textit{Kodak}/\textit{name of a company}, the underscore shows the \textit{wb} spanning silent interval) made no difference. The majority of the listeners judged this target to be a natural word. The silent interval before post-boundary /k/ was obviously tolerated as a normal stop closure time.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.1.png}
\caption{Targets with silent intervals at the \textit{wb}. Mergers are named on the x-axis. The duration of the silent intervals at the merger’s \textit{wb} is given in ms and indicated through a vertical bar which corresponds to the left y-axis. On top of each bar, the duration is once more specified. The graph shows, how many listeners (in per cent) have identified the respective merger correctly; it corresponds to the right x-axis.}
\end{figure}

The process of merger identification was of special interest in the case of dialect speakers. The hypothesis had been that listeners who were dialect speakers themselves, would tolerate dialectically coloured targets due to acoustic similarities to their own dialects. Four listeners from Saxony and 7 listeners from Baden and Swabia belonged to the dialect group. All 11 were compared to the other participants with \textit{SG} or other dialects (n = 39) with respect to the identification of mergers that showed \textit{[+ fortis]} stop lenition. For the other analysis regarding the identification of mergers with non-standard stress, the 7 listeners with Baden and Swabian dialects were compared to the other listeners
The dataset for the [+ fortis] stop lenition tolerance analysis comprised only word pairs with a merger constituent showing [+ fortis] stop lenition across the \[\text{wb}\] (e.g. \([\text{das Handwerk}]\text{k-lebt/craftsmanship is alive} \rightarrow \text{glebt/klebt/sticks, 3. p. sg.}\), 200 items in total. Accordingly, the stress analysis contained word pairs with non-standard stress in the merged constituent (e.g. \([\text{dass Wind und Wasser}]\text{wirtschaft[lich sind]/that wind and water are profitable, standard stress: 'Wasserwirtschaft/water economics}\), also 200 items in total. We applied a binominal logistic regression with boundary condition, dialectal affiliation and lexical frequency as fixed variables, recognition correct (binary: yes/no) as the dependent variable, participants and items as crossed random factors. No difference in the identification of targets with lenised initial stops was found between listeners with Saxon, Swabian and Baden dialects and all other listeners \((p > 0.6)\). There was a main effect of lexical frequency \((p = 0.0013)\). Concerning the identification of targets with “wrong” stress, listeners with Swabian and Baden dialects achieved the same results like the other listeners. No effect for dialectal affiliation was found \((p > 0.1)\). Again, lexical frequency showed a main effect \((p = 0.0051)\). We also pursued the question, whether the perception of targets depended on the speakers’ origin. The speaker group comprised 3 speakers of SG and one dialect speaker from Saxony. A binominal logistic regression with speaker origin as the fixed factor, recognition correct (binary: yes/no) as the dependent variable, participants and items as crossed random factors showed no effect for speakers’ origin \((p > 0.5)\).

### 4.2.4. Discussion

Experiment 1 showed that listeners identified a merger more confidently if it carried two or three \[\text{wb}\] markers instead of only one \[\text{wb}\] marker. Interestingly, two and three marker targets had similar identification rates - the presence of a third \[\text{wb}\] marker did thus not improve recognition further. We assume that the individual strength of a marker to signal a \[\text{wb}\] and the possible combinations of strong and weak markers played a crucial role in the process of \[\text{wb}\] identification. The issue of individual marker strength will be addressed in experiment 2. An analysis of marker combinations, however, was beyond the scope of this thesis.

One \[\text{wb}\] marker that was examined specifically was the silent interval. The analysis showed that the durations of silent intervals in the merged targets were too short to support merger identification. Mergers had been created by cutting two neighbouring
words or word parts out of the audio stream, so that a possible silent interval was naturally included into the merged word. Consequently, the presented stimuli contained silent intervals that are natural between words in IP-medial position in running speech. These intervals were tolerated as natural silence. Even the longest silent interval in a merger (113 ms, target: *Koda_k*) did not improve recognition (identification rate: 44 per cent). For the majority of the listeners, the silent interval did not render this target into sounding unnatural; they presumably tolerated the silent interval as a normal closure time for the final stop /k/. On the other hand, targets with very short silent intervals at the merging area (< 50 ms) were recognised confidently, probably due to the presence of another, more prominent wb marker. The literature discussed in section 2.4 supposes that silent intervals will become strong wb markers and distinguishable from stop closures when they exceed a crucial duration. This topic will be pursued in experiment 2.

Apart from silent intervals, dialect specific tolerances were investigated. The analysis suggests that dialectal issues did apparently not play a role. Since we had only a few participants per dialect, this result has to be interpreted within the limits of this experiment. In the perception of targets which showed phonetic specifics similar to their own dialects, listeners from Saxony, Swabia and Baden did not perform differently from the other listeners. We tested the tolerance of (1) word-initial lenition of [+ fortis] stops (natural in Saxon, Baden and Swabian dialects) and (2) words with non-standard stress (occurs in Baden and Swabian dialects). In the case of [+ fortis] stop lenition, merging had produced several items with an artificial dialectal colouring due to an assimilation effect across the wb relieving the stop of its feature [+ fortis]. In the case of non-standard stress, targets with interchanged primary and secondary word stress were analysed. The reason why the examined dialect group did not respond differently from the other listeners might be explained in terms of linguistic competence as described by Clopper and Pisoni (2005) (section 2.9). Each speaker who delivered acoustic stimuli in the perception experiment had been heard en-bloc. The three SG speakers came first, the dialect speaker last. The majority of the samples with supposed dialectal nuances had been produced by SG speakers and they were played when listeners had already heard samples which had been uttered in SG by the respective speaker. Presumably, after the perception of the SG stimuli, listeners knew that the person who delivered the targets was not a dialect speaker and they formed expectations on how the next target should sound. The sudden appearance of a dialectally coloured item has presumably prompted test persons independent of their dialectal origin to decide against the nat-
ural word. Accordingly, the stimuli delivered by the Saxon speaker had no impact on recognition either. We assume that listeners quickly identified the speaker as a person with a dialect and consequently altered their expectations towards the articulation of the coming words. The data from our perception experiment suggest that dialect was no interfering factor. However, this result needs to be verified in a larger study with more participants per dialect.

In the following, individual markers and their importance for boundary perception were put into focus. The question when silent intervals are long enough to function as markers was still open and thus pursued further. The crucial duration had not yet been reached. Dialect specifics were no longer an issue in the second part, because no effect had appeared in the first experiment. The number of participants was too small to get more than a first impression and the group of listeners with dialects in the second experiment was not larger.

4.3. Experiment 2

Perception experiment 1 had shown that merged targets with two or three markers were better recognised than mergers carrying only one marker. In the second perception experiment, all markers were examined separately according to their individual boundary signalling strength. The aim was to rank the markers from the strongest to the weakest. Another recording was made in order to get a larger number of targets for each marker. While in the first experimental part, the examination of dialect specifics required more than one speaker of the stimuli, the second experimental part needed only one. A speaker of SG was chosen to produce the acoustic samples, and he read the texts with the embedded word pairs twice. Two recordings allowed for the cross-splicing of our target words, the mergers. As a basic principle and considering that a merger consists of two parts which when joined together result in a meaningful word, the first part stemmed from the first recording, the second one from the second recording. Natural words were spliced accordingly; the identity-spliced portions corresponded to those of the respective merger. Sometimes, further manipulation became necessary with regard to the mergers. If they carried more than the one marker of interest, phonemes carrying additional markers were substituted by their correspondents from the natural word or erased in case the merger carried more phonemes than the natural word. The splicing/manipulation procedures should minimise the effects of recording (e.g. speech
4. Perception experiment

tempo) as well as ensure that only one [wb] marker remained within the merged target. Some alterations were made in the experimental design. Instead of a self-paced test, a speeded reaction test was chosen in order to get an immediate response and to measure reaction times. Data about our test persons’ quickness of response would allow us to draw conclusions about [wb] marker strength. If a merger, for instance, is predominantly judged as “unnatural” and reaction times are high on average, the [wb] marker is likely to be strong. Listeners delay their response, when they hear that something in a word is “wrong”.

While the main analyses centred around the question of [wb] marker strength, one additional check was made regarding merger identification according to construction (lexeme + lexeme versus lexeme + phoneme/phoneme sequence).

4.3.1. New Recording

Before we come to the actual perception experiment, the recording of the new reading material is explained.

Speaker
One male speaker, 39, produced the items for the second perception test. He spoke [SG] (origin: Mecklenburg-West Pomerania) and was neither a professional speaker nor had he a linguistic background. The speaker was chosen out of five candidates, because he delivered the acoustic samples with a very clear articulation. He had not been involved in the recordings for the production experiment and had not participated as a listener in the first perception experiment.

Recording materials
One hundred and sixteen individual carrier sentences were created that contained 58 word pairs, with a minimum of five pairs per marker. The examined [wb] markers were glottal stops, stop allophones, /l/-allophones, non-standard stress and silent intervals (see Table 4.1). Eleven of the word pairs from the previous experiment were re-used for the new reading material, all others were new (for the reading text, see Appendix A.1.5, for a target list, see Appendix A.2.1.2). Similar to the first experiment, a word pair consisted of a natural word and its merged twin (Monat, Mona_T). A carrier sentence contained either the natural word or the sequence for the future merger, never both. Constituents of a word pair had the same position within their respective sentences.
4. Perception experiment

(e.g. both word-initial) to control for prosody. The order of the sentences had been pseudo-randomised with Excel. Sentences were numbered with the purpose to provide orientation for the speaker. For the recording, an Audio Technica AT4050 SM microphone and a PCMCIA Foster FR-2 (48 kHz/24 Bit) recorder were used.

Recording procedure

The recording was made in an anechoic room of the recording studio in the University of Music and Theatre Rostock. The speaker read in a standing position, with the microphone hanging from the ceiling. A printout of the text was displayed on a note stand in front of him. The speaker was instructed to read the sentences aloud with a declarative intonation contour. This was practiced prior to the recording. During the recordings, the experimenter followed the session in another room via headphones with the reading material and a pen in front of her. Targets which had not been produced with the sufficient quality, were noted down (e.g. swallowed syllables, muddled words). After the recording, the experimenter requested the speaker to repeat the sentences containing these targets until the required standard (clear SG articulation) was reached. After that, another recording of all 116 sentences was made which also included the correction procedure as described. The acoustic material was reprocessed by a sound engineer in order to erase occasional side noise that was not immediately noticed during recording (e.g. mild feet shuffling).

The next section presents the methods and results of perception experiment 2.

4.3.2. Methods

4.3.2.1. Participants

Forty test persons took part in the second perception experiment (22 female and 18 male), aged between 21 and 56. Seventeen spoke dialect, all others SG. Thirty-three speakers had no linguistic background, 7 participants had taken speech training lessons.

---

8 We used the Excel RANDBETWEEN function for this. This function generates random numbers in a pre-defined range, in our case with 1 as the lowest number and 116 as the highest (total number of stimuli to be randomised). All numbers within this range are unique. We listed our natural and merged stimuli (representatives of the carrier sentences) in an Excel spreadsheet, one per line, and allocated a random number between 1 and 116 for each. Afterwards, stimuli were sorted according to the allocated number, starting with 1, ending with 116. The order was corrected manually when the two constituents of a word pair appeared as neighbours, which we wanted to avoid.

9 We thank Sebastian Paschen.

---

100
(elocution classes) as part of their studies; they were students and teachers at the University of Music and Theatre Rostock with majors/subject areas in Vocals, Lehramt Music or Drama in Education. Thirteen listeners were acquainted with the speaker of the acoustic stimuli. They had been asked after the perception experiment whether they had recognised the voice. One participant could name the speaker, all others had not recognised him.

4. Perception experiment

4.3.2.2. Materials

Natural and merged words from the new recording (section 4.3.1) were cut and spliced with Praat. The procedure is explained in the following using the example word pair \textit{Mona\textunderscore T[ölle]/first and family name - Monat/month}.

Merged word: \textit{Mona\textunderscore T[ölle]}
Natural word: \textit{Monat}

Two realisations for each sequence, one from recording 1 (r1), one from recording 2 (r2)

Cross-splicing of the merger: [mona], the name, from r1 + [th], the aspirated /t/ coming from word-initial \textit{Tölle} from r2
Identity-splicing of the natural pendant: sequence [mona] taken from \textit{Monat} from r1 + [t] coming from \textit{Monat} from r2

So far, merger \textit{Mona\textunderscore T} [monat\textsuperscript{h}] still carried two \textsuperscript{wb} markers. First, the stop allophone /t/ - the marker of interest - but the sequence also differed from the natural word with regard to quality and quantity of the vowel /a/. Consequently, the open syllable vowel in the target \textit{Mona\textunderscore T} had to be replaced by the closed syllable vowel coming from the natural word \textit{Monat}, using the same recording as the sequence that is manipulated stemmed from (here: r1).

Hence, the merger \textit{Mona\textunderscore T} has two splicing locations - first, [th] was spliced across mergers and second, [a] was spliced from the natural word into the merger.

The necessity to exchange phonemes due to unwanted quality/quantity differences between the natural and the merged word also applied to consonants.

The merger \textit{[auf dem Ti]sch\textunderscore Tiere/animals on the table} for instance, had been created to examine the impact of allophonic variations of stops. A highly aspirated stop was placed in a fricative-consonant cluster, where the stop is little or not aspirated.
### Table 4.1. Analysed wb markers, phoneme splicing and elimination procedures.

<table>
<thead>
<tr>
<th>Wb marker</th>
<th>Natural word</th>
<th>Merger</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal stop</td>
<td>kehren (to sweep)</td>
<td>[dan]k_ehre[amtlicher Helfer]/thanks to volunteers</td>
<td>(1) The speaker had produced all targets of this group with glottal stop only and not in combination with creaky voice. (2) The initial stop of the merger was substituted by the stop of the natural word.</td>
</tr>
<tr>
<td>Stop allophone: initial stop in final position</td>
<td>Monat (month)</td>
<td>Mona_T[ölle]/first and family name</td>
<td>A vowel before the wb was exchanged by the respective vowel of the natural word.</td>
</tr>
<tr>
<td>Stop allophone: final stop in initial position</td>
<td>Knoten (knot)</td>
<td>[wie star]k_Noten [voneinander abweichen]/how greatly musical notes differ</td>
<td>The phoneme after the wb was replaced by the phoneme of the natural word.</td>
</tr>
<tr>
<td>Stop allophone: initial stop in medial position</td>
<td>Skalp (scalp)</td>
<td>[er]stes Kalb/first calf, stop /b/ is uttered as [+ fortis] stop /p/ due to German final devoicing</td>
<td>The initial fricative of the merger was substituted by the fricative of the natural word.</td>
</tr>
<tr>
<td>/l/ allophone</td>
<td>Pappel (poplar)</td>
<td>Pappe_[l]iegt/cardboard lies</td>
<td>All pre-boundary merger constituents ended on schwa which was eliminated during the digital concatenation of word-initial /l/ taken from the right neighbouring word.</td>
</tr>
<tr>
<td>Non-standard stress</td>
<td>Busfahrplan (bus time schedule)</td>
<td>[verkehrt der] Bus_fahrplan[mäßig]/the bus goes according to schedule</td>
<td>All targets were compounds. The impression of stress shift was achieved due to an interaction of word and sentence stress. The principle is explained in section 4.2.1.2.</td>
</tr>
<tr>
<td>Silent interval</td>
<td>Vater (father)</td>
<td>Va#ter</td>
<td>The hash shows the place of the silent interval. It was artificially spliced into the target.</td>
</tr>
</tbody>
</table>

(natural word: *Stiere/bulls* [ʃtɪrə]). The duration of the fricative preceding the aspirated stop in this artificial consonant cluster, however, had to be similar the corresponding fricative’s duration in the natural cluster. Hence, the fricative coming from the natural cluster was spliced before the stop in the merged twin. Apart from quality/quantity
issues, the deletion of schwa in the case of the /l/-allophone word pairs was another manipulation done with the aim to ensure that the natural and the merged word only differed in one \textit{WB} marker. Merger \textit{Pappe}[^l]iegt]/cardboard lies \([p^b\text{ap}o\_l]\), for instance, differed from the natural word \textit{Pappel}/poplar \([p^b\text{apl}]\) in two ways - it contained a different /l/-allophone, the marker of interest, but also a schwa, which is not part of the natural word. Hence, schwa had to be eliminated in the merger. For a summary of the splicing/manipulation procedures, see Table \ref{tab:splicing}.

Silent intervals were artificially inserted into a word. In experiment 1, the longest naturally produced silent interval had a duration of 113 ms. We could show that the merger containing this silent interval had not been recognised as such, but was perceived as a natural word. Taking the measure of 113 ms as the lower limit, longer durations were inserted, starting at 110 ms, going upwards in 10 ms-steps up to 180 ms. For each duration, another target was chosen due to the demands of the experimental layout, which allowed a sequence - either the natural word or the respective merger - to be only once present in the stimuli set that was played to a listener.

Targets differed in the number of phonemes before and after the silent interval. Beside that, the silent period was either inserted after a vowel or after a consonant. Silent intervals targets and their specifics are listed in Table \ref{tab:silent-intervals}. This experimental layout allowed us to get a general idea about the boundary-marking capacity of silent intervals under different conditions. For a more systematic study, each target would have had to be checked for each silent interval duration. Such a procedure, however, which was beyond the scope of this thesis.

Mergers with stop allophones were created in such a way that they disobeyed the rules of stop aspiration in German speech. According to Kohler \cite{1995}, \textit{[+ fortis]} stop allophones in German can be aspirated in most word positions. On the other hand, the duration of the aspiration is controlled by the position of the stressed vowel. If the stop is located before the stressed vowel, its aspiration is longer than in post-vocalic position (ibid.). Furthermore, a \textit{[+ fortis]} stop is less or not at all aspirated if it is part of a fricative-stop cluster, a stop-lateral cluster, or a stop-nasal cluster. Our stop allophone mergers disobeyed these principles; they had the following characteristics:

1. Word position: Initial stop allophones were merged to a final position and final ones to an initial position. Through that procedure, items with a high degree of aspiration were placed in positions where stops with a lower degree of aspiration are
common and vice versa. Initial (highly aspirated) stop allophones were also merged into clusters where stops normally show little or no aspiration e.g. [ʃtʰiro] taken from the sequence [auf dem Tisch] taken from the sequence [auf dem Tisch]sch_Tiere/animals on the table. This position is referred to as word-medial position. The phoneme carrying the [wb] marker does not start the artificial word but is preceded by another phoneme or string of phonemes.

2. Position relative to the stressed vowel: Here, the group of initial stop allophones in final position is of special interest. In the mergers, e.g. zu_kommen/to come (natural word: Zug/train) or Mona_Tölle/first name and family name (natural word: Monat/month), initial stop allophones were placed after a primary stressed or secondary stressed vowel. Such a sequence - strongly aspirated stop after the vowel - should exaggerate the “wrong” allophone impression.

Note that during the merging procedure the same across wb assimilation phenomenon as reported for the production and for the first perception experiment showed up for the targets with final stop allophones in initial position, albeit produced by other speakers. For instance, when the target word k_Noten (natural word: Knoten/knots) was cut out of the context [wie star]k_Noten [voneinander abweichen]/how greatly musical notes differ, the merged word did not start with a [k], but with a lenised stop as the result of nasal assimilation; the resulting string was [gnoνn].

<table>
<thead>
<tr>
<th>Target</th>
<th>Si duration</th>
<th>No. of phonems</th>
<th>Inserted after V or C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sei#te/page</td>
<td>110</td>
<td>5</td>
<td>V</td>
</tr>
<tr>
<td>La#tein/Latin</td>
<td>120</td>
<td>6</td>
<td>V</td>
</tr>
<tr>
<td>mer#kte/realised</td>
<td>130</td>
<td>6</td>
<td>V</td>
</tr>
<tr>
<td>Va#ter/father</td>
<td>140</td>
<td>4</td>
<td>V</td>
</tr>
<tr>
<td>Rot#barsch/redfish</td>
<td>150</td>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>Des#pot/despot</td>
<td>160</td>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>Fuß#ball/football</td>
<td>170</td>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>er#klomm/climbed</td>
<td>180</td>
<td>6</td>
<td>V [y]</td>
</tr>
</tbody>
</table>

Table 4.2. Silent interval (si) targets and their specifics. The left column shows the place, where the silent interval was inserted as indicated by the hash. The inserted silent interval duration is given in milliseconds.

10 Apart from the assimilation effect across the wb, the stimulus had been produced with a glottal stop substituting word-medial stop and schwa.
4.3.2.3. Procedure

The perception experiment took place in quiet rooms. A Samsung NC10 netbook and Sony MDR CD 550 headphones were used. Similar to the first experiment, two word sets were created, each containing 58 acoustic stimuli: 29 natural words and 29 mergers in a pseudo-randomised order. Pseudo-randomisation was again done with Excel as explained in section 4.3.1. The constituents of a word pair never appeared in the same set. The natural word belonged to set one, the merged twin to set two or vice versa. Targets were played to the listeners and visualised on the netbook screen using DMDX (Forster and Forster, 2003). Test persons were tested individually. They sat in front of the netbook wearing headphones which were plugged into the netbook. The volume had been adjusted at a comfortable level. Participants were orally instructed before the experiment. A training session was carried out with 10 items prior to the actual experiment in order to familiarise participants with the voice of the speaker and the experimental routine. Training items comprised both merged and natural words; they were not played in the following experiment. After the training set, the experimenter answered questions - if any had arisen - and left the room. When the participant pressed the space bar, the oral instructions were once more summarised in writing on the computer screen. The second pressing of the space bar started the actual experiment. A beep of 200 ms duration announced the item 500 ms before its appearance. In this speeded reaction test, listeners saw a target on the netbook screen and simultaneously heard it via headphones. It was always the natural word that appeared on the computer screen, written in standard orthography, nouns beginning with a capital letter, adjectives and verbs beginning with a lower case character. The listeners’ task was to decide if the perceived word sounded natural. For “YES”, the AltGr-key had to be pressed with the right hand, for “NO”, the Alt-key with the left hand; left-handers had a reversed button set-up. Reaction times were measured relative to the offset of the acoustic stimulus. The choice had to be made within two seconds after the end of the acoustic stimulus. If the computer received no input, a text line saying “zu langsam” (too slow) appeared. Three-hundred ms after the time-out or after a key press, the beep signalled the next sample. Reaction times as well as input information (target judged natural/unnatural) were stored in a textfile.
4. Perception experiment

4.3.3. Hypotheses and research questions

Glottal stops

It was assumed that vowel-initial glottal stops should give strong cues about word boundaries. Samples like [Dan]k\_ehren\_amtlicher\_[k\_?er\_n] (natural word: *kehren*/*to sweep* [k\_er\_n]) should be judged as sounding unnatural.\(^{11}\) This hypothesis is in accordance with Nakatani and Dukes (1977) who demonstrated that glottal stops and creaky voice are important cues in the word recognition process. Despite the fact that we could not test this marker combination, we expected that glottal stops have an effective \(\text{wb}\) signalling function, even as singular markers.

The glottal stop as a schwa-substitute between stop and /n/ in the word-medial, morpheme-initial position ("[k\_\œn\_n]"), which had been regarded in the analysis of the production data, was not followed-up in the perception experiment. The fact that there is a second frequent place of appearance for glottal stops does not put their value as boundary signals into question. Morpheme-initial glottal stops occur in a very restricted neighbourhood, which allows easy detection and categorisation as a non-boundary place.

Stop allophones

The production experiment had revealed that intensity values of stop allophones did not differ as a function of their position in a word and distance to the stressed syllable. However, we expected word-initial stop allophones in final position, e.g. *[in dieser] Bar\_t\[tummeln sich]/in this bar cavort* (natural word: *Bart*/beard), to produce an effect of condition. The merged targets gave an audible impression of hyperarticulation. This had been achieved through the merging of an initial stop, which had originally been located before a stressed vowel, to a final position.

A word-final stop merged to the beginning of a word would also result in an effect for condition, but for other reasons. Assimilation effects across the \(\text{wb}\) had re-appeared in the new production data, e.g. *[wie star\_k\_Noten \(\rightarrow \text{[gno\_n]}*]. All of our targets were affected to different degrees. We therefore expected that listeners would evaluate them as sounding unnatural.

\(^{11}\) The stop in the merger was substituted by the stop of the natural word with the aim to avoid multiple \(\text{wb}\) markers.
As a third allophone location, initial allophones merged to a medial position were considered. Half of our targets in this category started with the fricative-stop cluster /f/ + stop (\(\text{\textit{auf dem Tisch}}/\text{animals on the table, natural word: Stiere/bulls}\)), the other half had the initial fricative-stop-cluster /s/ + /k/ (\(\text{\textit{eigene}}/\text{s_Konto/one’s own bank account, natural word: Skonto/discount}\)). Kohler (1995) explained that German stops /p, t, k/ are very little if at all aspirated in fricative-stop-clusters. If, however, a fully released and aspirated stop appeared in that position, the resulting sequence was supposed to sound unnatural to the listeners.

**L-allophones**

Allophonic variations of /l/ at unfitting phonetic positions within a word would support listeners to judge the word as “unnatural”. Depending on its neighbourhood, approximant /l/ has changing phonetic qualities and one /l/ can differ greatly from the other (see section 2.3.1). The phonetic properties of /l/ in Pappel/poplar, for instance, differ from those of /l/ in Pappe-liegt/\textit{cardboard lies}. The merger had been created through the concatenation of the word Pappe/cardboard and initial /l/ coming from the right neighbour word liegt/lie + 3rd p. sg, where it is uttered as a palatalised consonant. If such a palatalised liquid appeared in word-final position, where a non-palatalised one is the proper variant, listeners would judge the item as sounding unnatural. This identification scheme was expected to work for the whole /l/-allophone group.

**Non-standard stress**

Non-standard stress in compounds (e.g. \(\text{\textit{dass in der Feuerwehr}}/\text{\textit{Männer}}/\text{\textit{arbeiten}}/\text{men work in the fire brigade, regular stress: }\text{\textit{Feuerwehrmänner}}/\text{firemen}\)) was assumed to produce an effect. Test persons heard the targets in isolation and would therefore expect them to carry regular stress. Naturally, non-regular stress patterns are possible in certain contexts. However, context was not provided and listeners had no time to construct it since the speeded reaction test prompted an immediate response.

**Silent intervals**

We expected that the crucial duration for a silent interval’s capacity as a wb marker would be reached, because the experiment involved targets carrying relatively long silent
intervals (max. duration 180 ms). More precisely, we predicted that the critical duration was just above the longest silent interval tested so far which was 113 ms and had not supported \texttt{wb} identification in perception experiment 1. However, the Dutch listeners of de Pijper and Sanderman (1994) had judged silent intervals \textgreater 100 ms as an audible boundary in their native language and we expected that the minimal duration for boundary marking silent intervals in German would not be too far away from that value. We also assumed that durational ambiguity between a deliberately inserted silent interval and a tolerable stop closure duration would become distinguishable. Samples with a stop following after the silent interval (e.g. \texttt{La\#tein}) would give an answer to this question. Once the necessary duration was reached, we hypothesised an effect of condition, independent of the vowel-consonant-environment and the number of phonemes before and after the silent interval. Furthermore, we hypothesised that unnaturalness judgements would increase with growing silence duration.

**Phonemically identical words without \texttt{wb} markers**

In the production experiment, no differentiation pattern for phonemically identical words without \texttt{wb} markers was found (see section 3.3.4). They thus presented an ideal control group for the statistical calculations. Therefore, the data material also comprised such items. Out of the recorded word pairs, five were suitable as controls, e.g. \texttt{Holunderblüte}/\texttt{elder bloom} - \texttt{Holunderblühte}/\texttt{elder blossomed}. They had been constructed similar to the word pairs with non-standard stress, but had been uttered without stress opposition. In order to evaluate their usability, three naive listeners who had not been involved in any of the experiments, had been asked to judge likeness of mergers compared to their natural counterparts. The question they had to answer was: Do you hear the same word twice? Beside the five word pairs in question, ten other word pairs were played. The three listeners judged the five word pairs that were later used as the control group in the statistical analysis as consisting of two identical words.

One additional analysis was carried out that was concerned with merger composition (two merged lexemes vs. one lexeme plus phoneme/phoneme sequence).
Mergers with two lexemes vs. mergers with one lexeme + phoneme/phoneme sequence

We also pursued the question whether recognition depended on the phonemic structure of the merged targets. They showed two patterns: (1) one lexeme at each side of the wb (e.g. Jugend_leben; youth + to live, natural word: Jugendleben/youth life) or (2) a lexeme at one side of the wb and a non-meaningful sequence at the other (e.g. k_Noten; /k/ + Noten/musical notes, natural word Knoten/knots). We assumed that construction (1) was more often judged to sound natural.

Main research questions:

1. How do the examined markers differ with respect to their individual wb signalling strength?

2. What is the minimum silent interval duration so that silent intervals can function as wb markers?

3. Does an increasing silent interval go hand-in-hand with an increasing number of unnaturalness judgements?

For the additional analysis, the following research question arose:

Does merger composition influence recognition?

4.3.4. Results

In total, 2320 targets were analysed - 1160 word pairs, each consisting of a natural word and the merged counterpart. Twenty-six timeouts had been recorded. They could neither be explicitly allocated to the natural nor the merger group, nor to one particular marker. Timeout targets were considered as incorrect responses.

Lexical frequency was coded in frequency classes given by Wortschatz Universität Leipzig as before (for more details, see section 3.4.5) but not controlled. The analysed data spanned frequency classes 7-21 while the representatives of frequency class 21 with extremely rare words were the two compounds Holunderblüte/elderflower and Jugendleben/ youth life.
Four. Perception experiment

Seven listeners had taken speech training lessons as part of their education. We found that they performed the word naturalness judgement task better than the other listeners. Participants with speech training judged mergers more often as “unnatural” than participants without a linguistic background. A more detailed study, however, was not possible. The two listener groups were too unbalanced in numbers.

Like in the first perception experiment, analysis was based on binominal logistic regression. The method allowed us to test the interactions between two and more factors (categorical and continuous). For the individual marker test, only the merger group was examined. Marker type and lexical frequency were the fixed factors, naturalness judgement (binary: yes/no) the dependent variable, participants and items were the crossed random factors. There was no interaction between marker type and lexical frequency and no main effect of lexical frequency, so lexical frequency was removed as a factor from the logistic regression model. Merger composition was the control factor.

The identification of phonemically identical targets without markers was tested first to evaluate their potential as controls. Since we expected that a merger like Holunderblühte/elder blossomed and the corresponding natural word Holunderblüte/elder bloom would both be judged as sounding natural, we chose to contrast this group to targets with a marker that was also unlikely to produce an effect. The choice fell on the items with non-standard stress, because statistical pre-calculations suggested that the signalling capacity of this group was low. As predicted, no effect was found (p > 0.1).

As planned, we used phonemically identical mergers without markers henceforth as the control targets. Compared to this control group, the following markers improved recognition:

- Glottal stops, e.g. [dan]kehren[amtlicher Helfer]/thanks to volunteers (natural word: kehren/sweep, v.) - beta = 4.298, z = 5.251, p < 0.0001
- Silent intervals >= 120 ms, e.g. Fuß#ball - Fußball/football (beta = 4.107, z = 5.660, p < 0.0001)
- Initial stop allophones in medial position, e.g. [auf dem Ti]sch[Tiere/animals on the table (natural word: Stiere/bulls) - beta = 2.201, z = 3.081, p = 0.002
- /l/-allophones, e.g. Pappe[liegt]/cardboard lies (natural word: Pappel/poplar) - beta = 2.146, z = 2.942, p = 0.003
- Final stop allophones in initial position, e.g. [wie star]Noten [voneinander abweichen]/how greatly musical notes differ (natural word: Knoten/knot) - beta = 1.507,
There was neither an effect for initial stop allophones in final position (p = 0.254) nor for non-standard stress, as mentioned before.

Silent intervals were examined into greater detail. The aim was to get a general idea about at what duration the wb signalling function started and if an increasing silent interval duration resulted in a proportionally increasing number of unnaturalness judgments. We calculated a binomial logistic regression with silent interval duration as the fixed factor, recognition correct (binary: yes/no) as the dependent variable and participants and items as crossed random factors. Targets with durations $\geq$ 120 ms showed an effect for marker type when contrasted to the control group (p-values $< 0.002$) with the exception of the 150 ms-target Rot#barsch/redfish which we will discuss later. Results for the silent intervals targets are given in Table 4.3.

<table>
<thead>
<tr>
<th>Silent Interval</th>
<th>Beta</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>120 ms</td>
<td>2.7478</td>
<td>3.493</td>
<td>0.0005</td>
</tr>
<tr>
<td>130 ms</td>
<td>4.8419</td>
<td>3.567</td>
<td>0.0004</td>
</tr>
<tr>
<td>140 ms</td>
<td>3.7578</td>
<td>3.464</td>
<td>0.0005</td>
</tr>
<tr>
<td>150 ms</td>
<td>1.0791</td>
<td>1.280</td>
<td>0.2006</td>
</tr>
<tr>
<td>160 ms</td>
<td>3.2020</td>
<td>3.275</td>
<td>0.0011</td>
</tr>
<tr>
<td>170 ms</td>
<td>3.7578</td>
<td>3.464</td>
<td>0.0005</td>
</tr>
<tr>
<td>180 ms</td>
<td>4.5957</td>
<td>3.385</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

Table 4.3. Results of the binominal logistic regression model for silent intervals. Durations were compared to a silent interval of 110 ms.

A silent interval duration of 110 ms produced a negative effect compared to the next following 120 ms-interval, which was significant (beta = -2.748, z = -3.493 p = 0.0005). Figure 4.2 illustrates that the identification of mergers did not increase proportionally to the expanding durations of silent intervals. The step between 120 and 130 ms showed an improvement of recognition, afterwards scores remained high. This result is in accordance with de Pijper and Sanderman (1994) who reported that increasing silent interval durations did not differ in perceived boundary strength once a crucial duration is exceeded (see 2.4.2). At 150 ms, however, we found an outlier (target Rot#barsch/redfish);

\[ z = 2.103, \ p = 0.036 \]  

\[ ^{12} \] Interestingly, when the reference was a strong boundary marker, the picture changed. With glottal stops as controls, the only group that was equally strong was the silent interval group (p > 0.8), all other markers produced significant negative effects (all p-values $< 0.007$).

\[ ^{13} \] All participants had recognised target mer#kte (si = 130 ms) correctly. We introduced one fictive instance (recognition correct = no) in order to get interpretable results, cf. Braun and Chen (2010, p. 15).
recognition fell back to the chance level. *Rotbarsch* differed from the other targets in such a way that the silent interval was framed by two stops. When looking into articulatory detail, we found that stop /t/ was not yet fully released when the closure for /p/ started. We therefore assumed that listeners tolerated the silent interval as two successive stop closure durations.

The number of phonemes a merger containing a silent interval consisted of did not play a role in the identification task (p > 0.6), neither did the place after which the silent interval was spliced into - after a vowel or after a consonant (p > 0.7).

One aim of the perception experiment was to present a ranking list displaying the strength of each *wb* marker. Markers were ranked according to correct merger identification (=unnaturalness judgement). Figure 4.3 underlines the capacity of glottal stops and silent intervals >= 120 ms to signal a *wb*. Glottal stops were identified best, 84 per cent of the mergers carrying this *wb* were correctly recognised. Silent intervals >= 120 ms came second (74 per cent), initial stop allophones in medial positions third (63 per cent). All three markers had produced an effect in the statistical analysis. They were better identified than the control items. The ranking list continues with non-standard stress, listeners recognised 59 per cent of the targets. After that follow /l/-allophones (56 per cent), final stop allophones in initial position (53 per cent), and initial stop allophones in final position (49 per cent). While the last two also reflect the results of the statistical analysis (no effect of marker typer), non-standard stress had produced no effect while /l/-allophones had - but the ranking list shows non-standard stress before
/l/-allophones. The bar chart does not necessarily reflect all findings from the statistical calculations. This is due to (1) the different form of reference criterion, (2) the inclusion of more factors in the statistical analysis and (3) the relation of the statistical analysis to a freely-chosen control group.

![Identification of wb markers](image)

**Figure 4.3.** Ranking of markers according to their strength to signal a wb. The ranking criterion was “per cent recognition correct”.

**Mergers with two lexemes vs. mergers with one lexeme + phoneme/phoneme sequence**

Two merger types were used in the perception experiment. The first type had one lexeme at each side of the wb, the second type had a lexeme at one side of the wb and a non-meaningful sequence at the other. All mergers of the dataset (n = 1160) were coded according to their composition and analysed. A binominal logistic regression with merger composition as the fixed factor, recognition correct (binary: yes/no) as the dependent variable and participants and items as crossed random factors delivered no effect of merger composition (p > 0.1).

**Reaction times**

In the experiment, reaction time measurement started immediately after the offset of the acoustic stimulus. Participants had two seconds for a yes/no input on the keyboard.
In total, 27 timeouts occurred, which had counted as incorrect responses in the previous statistical calculation. In the analysis of reaction times, however, they were excluded.

A logistic regression model was used with reaction time as the dependent variable, item type (natural word/merged word) and marker type as fixed factors, participants and items as crossed random factors. Reaction times for natural words were significantly faster than for merged items (\( p = 0.005 \)).

In a next step, only the merger group was examined. Decisions for correctly and incorrectly identified items took equally long (\( p > 0.18 \)). It appeared that reaction times depended on the \text{wb} \text{ marker} present in the target. With phonemically identical items without \text{wb} \text{ markers} as the control group, all other markers prompted longer reaction times (wrong stress: \( p = 0.01 \), initial stop in final position: \( p = 0.001 \), all other markers: \( p < 0.0001 \)). The two strongest \text{wb} \text{ marker} types glottal stop and silent interval \( \geq 120 \) ms had the longest mean reaction times (1465 ms/1490 ms), followed by initial allophones in medial position (784 ms), /l/-allophones (770 ms), final allophones in initial position (536 ms), initial allophones in final position (408 ms), non-standard stress (302 ms). Surprisingly, the mean reaction time for non-standard stress mergers is very short.

Since these items were relatively long compounds, one might assume that they were evaluated before their offsets and the response was given before the word had been heard completely. In order to test this assumption, all mergers were coded according to their number of phonemes. Items with six phonemes or less were labelled as “short” (\( n = 829 \)), all others as “long” (\( n = 317 \)). We again applied a logistic regression, now with number of phonemes (binary: short/long) as the fixed factor, reaction time as the dependent variable, participants and items as crossed random factors. We found no effect for number of phonemes. Apparently, reaction times did not depend on the length of a target (but note that our dataset contained more short than long items).

### 4.3.5. Discussion

The aim of experiment 2 was to examine individual marker strength. For this purpose, each merger was manipulated in such a way that it contained only one \text{wb} \text{ marker}; the rest of the sequence was identical to the natural word with respect to number of phonemes, phoneme quality and quantity. The experiment was designed as a speeded...
reaction test in which participants had to make a naturalness judgment about acoustic stimuli. Targets were presented both acoustically and visually.

Five \textcolor{red}{wb} markers were examined: glottal stops, stop allophones (initial stop in final position, final stop in initial position, initial stop in medial position), /l/-allophones, non-standard stress and silent intervals. Silent intervals were between 110 ms to 180 ms long, differing in steps of 10 ms. Phonemically identical word pairs without \textcolor{red}{wb} markers were used as the control group. Targets with silent intervals $\geq 120$, glottal stops, stop allophones (final in initial position and initial in medial position) and /l/-allophones were more often judged as sounding unnatural than the control targets; the effect was significant. No effect was found for targets with initial stop allophones in final position and for targets with non-standard stress.

As hypothesised, mergers with \textbf{glottal stops} received high identification scores. The result confirmed the findings of Nakatani and Dukes (1977) for English. When glottal stops were put up as reference in the statistical evaluation, only \textit{silent intervals} proved to be of equal strength, all other boundary markers produced negative effects. This is in line with other authors who highlighted the potential boundary function of silent intervals, such as de Pijper and Sanderman (1994) or Warren and Obusek (1971).

While 120 ms seems to be a plausible minimum duration signalling a \textcolor{red}{wb}, a more detailed analysis is required to substantiate this claim. Since silent interval duration was manipulated between items, further research adding a within-items manipulation is needed to confirm our conclusion. After all, the used stimuli differed in their morphological composition, the voicing of consonants and their place of articulation after the silent interval, word stress and lexical frequency.

Mergers with glottal stops and those with silent intervals $\geq 120$ ms showed the longest reaction times; test persons needed 1465 ms (glottal stops) and 1490 ms (silent intervals) to make their decision. Targets with other \textcolor{red}{wb} markers had reaction times below 800 ms. The great step in reaction times between the two strong boundary markers and the others underline the effectiveness of glottal stops and silent intervals $\geq 120$ ms. Because the present boundary marker altered the canonical sound of the word, the acoustic stimulus could presumably not be mapped to the mental lexicon’s phonological representation of the visually provided target and decision-making was delayed.

In the case of \textbf{initial stop allophones in final position}, which were generally judged as sounding natural, our findings supported the position of Kohler (1995) who explained
that in German speech, both initial and final stops can be articulated with aspiration. The individual degree of a stop’s aspiration might have been within the limit of tolerable phonetic variation (cf. Gaskell and Marslen-Wilson, 1996). Our hypothesis after which initial stop allophones in final word position would sound hyperarticulated to the test persons - judging from auditory impression rather from measured data since intensity levels of final and initial stops showed no differences - was thus not verified. Measured and perceived data was in tune. One might also argue that the final position of the allophone had some effect on our results. Listeners might have have judged a target before its end, because for them the uniqueness point of the word had been reached earlier (Marslen-Wilson and Tyler, 1980).

**Initial stop allophones in medial position** produced the hypothesised effect. All our mergers started with the sequence fricative + aspirated stop allophone + vowel. In a fricative-stop-cluster, the stop is little or not at all aspirated in German (Kohler, 1995). Our stops, on the other hand, had the strongest possible aspirational cues in German coming originally from a word-initial position preceding a stressed vowel. Conceivably, the “wrong” allophone has prompted the effect. Also, the cue was available early which had probably a stronger influence on listeners’ perception than the previously discussed final allophone position.

For targets with **final stop allophones in initial position**, the expected effect occurred as well. Mergers carried a [- fortis] stop in a position where SG demands a [+ fortis] stop. Assimilation across the word boundary had produced that effect. Consequently, listeners did not accept these targets as natural words, when they were heard in isolation.

Our assumption that **non-standard stress** in compounds presented out of context would support merger recognition did not apply. An exchange of primary and secondary stress did not result in unnaturalness judgements. Listeners made their decisions fast. The mean reaction time was 302 ms. Apparently, listeners were quite confident that the perceived target was a natural word. Small, Simon, and Goldberg (1988) had shown that English listeners tolerated mis-stressed words if the resulting sequence was not a non-word. The verb *insert*, for instance, was acceptable if it had the intonation pattern of the noun *insert*. Obviously, this was also the case for our targets. Two examples shall illustrate that:

*Busfahrplan/bus time schedule* - Not any timetable (Fahrplan), but the bus time schedule vs. *Busfahrplan* - Not the system map of the bus, but the time schedule of the bus.
4. Perception experiment

‘Feuerwehrmänner/firemen - Not any men (Männer), but those of the fire brigade vs. Feuerwehr’männer - Not the firewomen, but the firemen.

The results also confirmed the findings of Domahs et al. (2008) who had explained that the tolerance of mis-stressed words depended on the position of the mis-stressed syllable. Stress violation may be tolerated if it does not demanded a re-parsing of syllables into feet. The conclusion that stress shift does not necessarily impair word recognition had also been drawn by Small, Simon, and Goldberg (1988). As long as the sequence is meaningful, non-standard stress is tolerable.

Our expectation concerning the [wb] marking strength of /l/-allophones was verified. The acoustic properties of /l/ depend on the neighbourhood of the liquid (Kohler, 1995). If the phoneme /l/ had improper phonetic features, the majority of the test persons judged the candidate as sounding unnatural. The mean reaction time of 770 ms was in the middle area of the obtained reaction times. Test persons’ decisions about /l/-allophones clearly took longer than decisions on targets which sounded natural to them, e.g. targets with non-standard stress (mean reaction time was 302 ms). This fact speaks for a relatively well-perceived [wb] marker.

Silent intervals were investigated in greater detail in order to answer the question: How long has a silent interval to be to function as a [wb] marker? Mergers became not distinguishable until the silent period spliced into the target was at least 120 ms long. A shorter silent interval had no effect. Another question we followed up was if the number of unnaturalness judgements increase proportionally to silence duration. This was not the case. Merger identification scores for items with silent interval durations >= 120 ms were high on a constant level, which pointed at a categorical perception of the merged targets. There was one exception. Target Rot#barsch/redfish with a silent interval of 150 ms received the same number of correct and incorrect responses (as before, the hash indicates the place, where the silent interval was spliced into). The only other important point in which Rot#barsch differed from the rest of the targets was that the item’s silent interval had been put between two stops. We therefore concluded that the silence of 150 ms was an acceptable time for two successive stop closure durations. The specific phonemic environment before and after the silent interval obviously influenced the test persons perception. As the last criterion, we examined if the number of phonemes or the location, after which the silent interval was inserted (vowel or consonant) had an effect, but we found none. Note that the minimum duration we suggest here as a [wb] marker represents a rough measure rather than a generalisable result. The used stimuli differed
4. Perception experiment

in several conditions (e.g. word stress, morphological composition). An in-depth analysis with each of the eight targets combined with all eight silent interval durations would have been necessary to provide a more detailed insight into the perception mechanisms. Such an examination was beyond the scope of this thesis.

One major aim of the second perception experiment was to rank the examined markers according to their individual \textbf{wb} signalling strength. The ranking list was assembled using “recognition correct in per cent” as the criterion. The three strongest \textbf{wb} markers were glottal stops which were identified best, followed by silent intervals $\geq 120$ ms and initial stop allophones in medial position (consonant clusters).

Apart from the impact of individual \textbf{wb} markers, one side factor in connection with word perception was tested, namely the influence of merger composition. Our prediction was that merger composition influenced the naturalness judgement, but no effect showed up.

The last chapter summarises the results of this thesis, evaluates the experimental conditions of the perception part and gives an outlook for future fields of research.
5. Summary of the results, evaluation of experimental conditions, outlook

This dissertation has provided new insights into the nature of \textit{wb} markers. They were discussed in reference to lexemes. A speech corpus was created that contained word pairs consisting of a natural word like \textit{Monat/month} and a phonemically identical merged twin \textit{Mona_T/"olle//first name and family name}. The merger was formed through the concatenation of neighbouring words or word parts, the underscore shows the \textit{wb}. Through that procedure, a \textit{wb} was inserted. A production experiment examined unpredictable \textit{wb} markers which were termed this way because it is uncertain whether a speaker uses them or not. These were glottal stops/creaky voice, allophonic variations of stops, silent intervals, contact geminates, and comparative vowel duration.

Depending on the \textit{wb} markers, we tested different characteristics: (1) place and frequency of occurrence (glottal stops/creaky voice, silent intervals), (2) differences in intensity levels (stop allophones), (3) duration (silent intervals, comparative vowel duration, contact geminates).

A two-part perception experiment investigated if \textit{wb} markers are perceptible. Experiment 1 was designed as a self-paced test; participants heard an acoustic stimulus (natural word or merger) and had to decide from which of two possible sentences the target originated. We followed up the question if listeners’ identification of a merger depended on the number of \textit{wb} markers present in a target. Experiment 2 analysed the individual strength of chosen \textit{wb} markers (glottal stops, allophonic variations of stops and lateral /l/, silent intervals, non-standard stress) in a speeded reaction test with isolated items. Acoustic stimuli again comprised natural and merged items, but this time, mergers carried only one \textit{wb} marker.

The results of the production and perception experiments are summarised in section 5.1. In order to focus on the phonetic component of word perception, other mechanisms linked
with speech recognition (e.g. syntactic segmentation) had to be suppressed. Hence, the word recognition task did not reflect the natural process of speech recognition. The issue will be addressed in section 5.2. Future research that may build upon the results of this thesis is outlined in section 5.3.

5.1. Experimental results

The production experiment with 19 participants showed that glottal stops were both frequent and reliable with regard to their two common places of occurrence. They usually appear word-initially if the word starts with a vowel and in a word-medial, morpheme-initial position as stop substitutions. For our analysis, the word-initial position was of special interest. In view of the fact that glottal stops often appear with creaky voice segments as their companions, creaky voice was also analysed in connection with the examination of glottal stops. Two-thirds of the targets with initial vowels were realised either with a glottal stop or with creaky voice or with both. The most common realisation was a creaky vowel onset. The frequent appearance of glottal stops and/or creaky voice in the word-initial position and the easy identifiability of glottal stops in the word-medial location recommended the marker/marker combination as a wb signal.

Another marker with a high potential to indicate a wb is the contact geminate. Exemplarily, nasal consonants were examined with regard to their temporal organisation. German has no geminates in its phonemic inventory. However, contact geminates may arise when one word ends and the following one begins with the same phoneme (e.g. einem Mann/one man, dat.). It was demonstrated that nasal contact geminates significantly differed in their durations from nasal singletons in word-initial, word-medial and word-final but non IP-final positions. Most IP-final singletons had undergone final lengthening; they showed equal or longer durations compared to contact geminates. A method to distinguish lengthened singletons from contact geminates in an automatic classification with segment duration as the criterion was presented. Identification became possible through the examination of the segment’s right neighbourhood. If the nasal was followed by a silent interval, the segment had to be an IP-final singleton. Phonotactic rules of German demand that nasal contact geminates have to be followed by a vowel. A silent interval after a contact geminate is not possible.

While glottal stops/creaky voice and contact geminates appeared to be useful wb sig-
5. Summary of the results, evaluation of experimental conditions, outlook

nals, other markers investigated in the production experiment implied only little or no potential at all. Among them were stop allophones. A supposable wb function of stop allophones may result from differences in aspiration. In German, however, stops may be aspirated in most positions of the word (Kohler, 1995), which our acoustic data verified. The measurement of aspirational cues was considered not to be the method of choice, because the raw values obtained in our production study were difficult to compare. Instead, we analysed intensity levels of stop allophones in relation to the nearest full vowel in different word positions as indicator for consonantal strength. The results showed no effect of position, but the standard deviation within the dataset was high.

Silent intervals were examined with the aim to test if they were longer in the merged word than in its natural counterpart. For the analysis, silent intervals at the wb area of mergers were measured and so was the corresponding area in the natural word. A longer silent interval in the merger might signal a wb since silent periods in fluently produced natural words are simply caused by the physical constraints of the vocal organs as opposed to deliberate insertion of silent intervals with the aim to divide two sequences. We examined the silent intervals in mergers and in the corresponding areas of the natural counterparts, e.g. merger auf#Bau [und ... achten]/be sure to pay attention to the build-up and... and natural word Auf#bau/build-up (the hash shows the place of the silent interval), but found no effect. Deliberately inserted silent intervals and those which are part of a phoneme such as stop closures were not distinguishable.

The findings of Rietveld (1980) for French speakers to discriminate phonemically identical sequences with differing syntax through comparative vowel duration could not be verified for German. Rietveld had found that sequences like [myska] in the noun phrase le muscat perdu/the spoiled muscat /wine/ and in the sentence le musc a perdu/the musk is evaporated, were distinguishable through the opposing duration patterns of the vowels, among other cues like the rhythmical structure of French. In the given example, the noun phrase had a short /u/ and a long /a/ while the sentence showed a long /u/ and a short /a/. Opposing vowel duration, however, could not be found in our German word pairs. They had been produced without any wb marker. Primary and secondary stressed vowels in the examined phonemically identical, but syntactically differing phrases showed no durational contrast, for instance diphthong /au/ and vowel /a/ in Augenmaß/sense of proportion versus [mit den] Augen_ maß/measured by the eye.

The perception experiment, which consisted of two parts, examined the role of wb
markers in the process of word recognition. Experiment 1 demonstrated that mergers with several \texttt{wb} markers were better identified than those carrying only one. The object of experiment 2 was the analysis of individual \texttt{wb} marker strength. Significant effects for a better merger identification were found if the target carried either a glottal stop, a silent interval \( \geq 120 \text{ ms}, \) an initial stop allophone in medial position (fricative-stop clusters), an /l/-allophone or a final stop allophone in initial position. No effect was found for mergers with initial stop allophones in final position or non-standard-stress. Phonemically identical mergers without \texttt{wb} markers - those word pairs which had previously been tested for comparative vowel duration - were used as the control group.

The individual results for each \texttt{wb} marker were analysed. Glottal stops fulfilled the expectations regarding their \texttt{wb} signalling strength - mergers were confidently recognised. Silent intervals, which had been artificially inserted into target words, showed an effect once they reached a minimum duration of 120 ms. After that point, listeners judged items as sounding unnatural. However, there was one exception with the item that carried a silent interval of 150 ms (\texttt{Rot#barsch/redfish}, the hash shows the location, where the silent interval was spliced into). The target received the same number of correct and incorrect responses. Only in this stimulus, two stops framed the \texttt{wb}. The pre-boundary stop was not yet fully released when the closure for /p/ started. Listeners obviously accepted this as a normal stop closure duration in a word. The example demonstrates that even relatively long silent intervals may be tolerated as being part of the normal articulation process. Stop closures times and (unmanipulated) silent intervals in running speech have approximately equal durations and can hardly be distinguished from one another. Since the targets we used for the analysis of silent intervals differed in several conditions, our results give a first approximate measure regarding the \texttt{wb} signalling function of silent intervals but do not allow for generalisation.

Mis-placed stop allophones in mergers were perceptible in some conditions. Initial stop allophones were merged to a medial location and became thus part of a fricative-stop-cluster that is hardly if at all aspirated in German. We assume that this acoustic cue was crucial for the listeners’ decision. The effect found for final stop allophones in initial position probably originated from an assimilation process across the word boundary which had produced targets with [- fortis] stops in positions where \texttt{SG} demands [+ fortis] stops. No effect was found for initial stop allophones in final position. Despite the highly aspirated stop at the end of the word, listeners perceived the target as natural.
This result reflects the findings from the production data: stop allophones, irrespectively of their position in a word, did not differ in intensity levels and thus provided no cues for the listeners. However, an alternative interpretation is conceivable: The cue came too late, and listeners might have made their decision before a target’s end had been reached.

There was also no effect of marker type for targets with non-standard stress, all of them were compounds. Stress shift did not render the merged targets into sounding unnatural. The exchange of primary and secondary stress in a compound is not unusual in German. Depending on the context, a subordination of the formerly dominating word is perfectly acceptable. For example, the word Busfahrplan/bus time schedule can be segmented in two ways - with primary stress on the first constituent as in “Not any time schedule, but the bus time schedule” or with primary stress on the second constituent as in “Not the system map of the bus, but the timetable of the bus”.

One major aim of the perception experiment was to rank the examined markers according to their strength. For this representation, the criterion “recognition correct in per cent” was chosen. Glottal stops were best identified, 84 per cent of the listeners judged mergers carrying this marker as sounding unnatural. Silent intervals >= 120 ms came second (74 per cent), followed by initial stop allophones in medial positions (63 per cent).

### 5.2. Evaluation of the experimental conditions in the perception tests

This work was concerned with phonetics. In order to isolate the phonetic part in the process of word recognition, other mechanisms that work in parallel, had to be suppressed. Consequently, the tasks performed by the participants of our perception studies did not reflect the normal speech recognition process. In both perception experiments the judgment task was focused on the isolated word which had been cut out of context.

Gaskell and Marslen-Wilson (1996) explained that “experiments on isolated words may capture only a subset of the properties of the speech recognition system, and where phonological variation is concerned, it may be necessary to study longer units of speech” (ibid., p. 146). Remez (2005) mentioned another negative effect of isolated perception studies. They are often carried out in a restricted domain where speech is the only
sound. A laboratory situation cannot mirror the challenges the perceptual system is exposed to in every day life.

While the criticism of the researchers cited above is acceptable in general, phonetic experiments have to preclude other speech recognition strategies in order to reach a specific goal. Generally, the perception of isolated targets in quiet surroundings was essential for our experiments. We have, however, attempted to account for factors which might negatively influence the perception of isolated targets. First, human voices were used instead of synthesised ones. Second, there had been training phases before every experiment in order to familiarise test persons with the (unusual) task ahead. None of our listeners had taken part in linguistic experiments before. Third, in order to avoid errors in the decoding of word meaning, targets were presented acoustically as well as visually so that listeners could concentrate on the naturalness judgment. A forth point is that lexical frequency of the targets, which is an important factor in word perception, was always considered in the statistical calculations.

One might argue that artefacts in the spliced mergers might have caused unnaturalness judgements rather than the markers we were interested in. The considerable number of naturalness judgements for mergers, however, shows that our acoustic stimuli were of an acceptable quality in general.

Apart from the naturalness of targets, the linguistic background of listeners may influence the results of phonetic experiments. Schouten, Gerrits, and Hessen (2003) observed a higher ability to discriminate in test persons with phonetic knowledge. Our observations confirm this finding. While our perception studies were generally conducted with linguistically naive listeners, seven participants of the second perception experiment had a linguistic background in the form of speech training. Their word judgement performance was better compared to that of the untrained test persons.

5.3. Future Research

The thesis presents results about the impact of selected markers on word segmentation. Future research might deal with their interaction. Which combinations are most effective? Some sections have briefly touched upon the topic. Glottal stops and creaky voice, silent intervals and segment lengthening were common combinations that appeared in our acoustic data. How effective are they from the perception point of view?
A [wb] marker interaction matrix might shed light on this question.

Our findings are also interesting for speech technology, especially the results about the temporal organisation of contact geminates. On-line measurement of consonantal duration might give cues for the location of a [wb] in machine-based speech recognition. Real geminates are not present in German speech, which means that exceeding a pre-defined phoneme duration may be used as a [wb] signal. We examined nasal segments which had been obtained from two separate production datasets. Nasal segments from both sets were best separable into geminates and singletons at 100 ms. Items below that duration were classified as singletons, items that were longer as geminates. However, this cannot be a fixed value, not even for nasals, since articulatory characteristics of speakers are highly individual. Consequently, the critical phoneme duration as a divisor of singletons and geminates will have to be re-adjusted for every phoneme and speaker or speaker group. Section 3.4.5 also suggested a procedure for the detection of lengthened singletons in IP-final conditions through the combined presence with IP-final silent intervals. This is an additional step that works with the purpose of error correction. It would be interesting to know if the mechanisms suggested in connection with the automatic classification of singleton and geminate items holds in machine-based speech recognition.

5.4. Zusammenfassung


Um die Untersuchungen durchführen zu können, wurde ein Sprachkorpus erstellt. Dafür wurden 19 Sprecher beim Lesen von Texten aufgenommen, neun von ihnen waren Dialektsprecher. In den Lesetexten waren spezifische Wortpaare eingebettet, welche aus einem natürlichen Wort, z. B. Monat, und einem künstlich zusammengesetzten Gegenstück, z. B. Mono_T bestanden. Das künstliche Wort wurde dabei zusammengesetzt aus...

Produktionsexperiment

Neben Plosiv-Allophonen wurden Glottalverschlüsse analysiert. Von Interesse war, an welchen Positionen und wie häufig sie auftreten. Parallel betrachtet wurden Knarrlaute,

---


---

126


Ein ähnlich hohes Potential zur Anzeige von Wortgrenzen wurde bei den stillen Inter-


Die Analysen von fünf Wortgrenzenmarkern im Rahmen des Produktionsexperiments zeigten, dass Glottalverschlüsse und Kontaktgeminaten zur Anzeige von Wortgrenzen

Perzeptionsexperiment

Worte so manipuliert, dass jedes nur einen Wortgrenzenmarker trug. Die statistischen Analyse der Kunstworte (lineare Regression, Kontrollgruppe: Kunstworte ohne Wortgrenzenmarker) zeigte Effekte für:


4. /l/-Allophone, z. B. Pappe_l - zusammengesetzt aus die Pappe liegt, natürliches

---

Die wortinitiale Aspiration von /k/ im Kunstwort wurde der des natürlichen Wortes angepasst.
5. Summary of the results, evaluation of experimental conditions, outlook

Wort: Pappel. Der Liquid befand sich jeweils in einer unpassenden Phonenumgebung, und die Kunstworte wurden vermutlich deshalb als solche identifiziert.


Insgesamt konnte in dieser Arbeit gezeigt werden, welche Potentiale Wortgrenzmarker als Grenzsingale haben. Die Präsenz von mehr als einem Marker verbessert die Erkennung einer Wortgrenze. Welche Marker-Kombination hierbei am effektivsten ist, könnte der Gegenstand weiterführender Forschung sein. Weiterhin wurde die große Stärke des Glottalverschlusses als Wortgrenzmarker hervorgehoben und eine Mindestdauer für stille Intervalle vorgeschlagen, ab welcher sie als Grenzsignal fungieren können. Von Interesse für thematisch angrenzende Arbeiten im Bereich der maschinellen Spracherkennung könnten die Funde in Bezug auf Kontaktgeminaten sein, die ebenfalls eine gute Verwendbarkeit als Wortgrenzmarker zeigten.


A. Appendix

Sections A.1.1 to A.1.4 show all reading texts used for the production experiment. All texts show the natural words and future mergers in their respective contexts. Targets are underlined. The first three readings texts (sections A.1.1 to A.1.3) also contain the targets used for the first analysis of geminate and singleton segments. They are shown in brackets and the measured nasals are given in bold letters. Sometimes, sequences have been used for the study of geminate/singleton segments as well as for the study of other wb markers, for example the sequence [einem Monat]. For the analysis of geminate/singleton durations, we have measured three segments (bold), namely /n/, /mm/ and another /n/. In addition to that, target Monat was also used for another wb marker test and is therefore underlined. Section A.1.4 shows the reading text used for the second analysis of geminate/singleton segments. The measured nasals are again represented with bold letters.

The underlined targets of readings texts A.1.1 to A.1.3 were not only used for the production experiment, but also for the perception experiment 1. For perception experiment 2, another recording was made, section A.1.5 shows the reading text; the underlined targets were later presented to listeners.

Section A.2 refers solely to the perception experiment. All targets are listed along with the information for what wb marker test(s) each item was used.
A.1. Reading Texts

A.1.1. Production experiment, text 1


Da gab es derbe Dialoge, aber auch freundschaftliche Gespräche. Schubert selbst sieht sich nicht als Störenfried, eher als Vorkämpfer für eine intakte Umwelt. Die Verkaufszahlen geben ihm Recht: Der Bekanntheitsgrad des Buches steigt.

Ob er Tipps dazu hat, welchen Beitrag jeder Einzelne zum Umweltschutz leisten kann? “Davon kann ich den Lesern genau acht geben, die ich auch einhalte”, sagt der Autor und zählt auf:

Erstens: Nicht bei laufendem Wasser Zähne putzen.
Zweitens: Tagsüber Heizung herunterdrehen.
Drittens: Auf Solarenergie umrüsten.
Fünftens: Alte Elektroleitungen erneuern.
Sechstens: Die Spülmaschine nur anstellen, wenn sie ganz voll ist.
Siebentens: Die Belüftung der Wohnung sicherstellen.
Und achtm: Eine Anlage zur Erdwärmegewinnung installieren.
Aber Vorsicht - das teure Gerät gut sichern, Diebe lauern überall.
A.1.2. Production experiment, text 2

Der unheimliche Baron


Der Baron blickte düster, während er Egon und Frank mit den Augen maß. „Sie müssen es gut haben, hier in ihrem Schloss“, versuchte Egon die Stimmung zu lockern. Der Baron kümmerte sich nicht darum. Er führte seine Gäste durch das Schloss.

Egon sah dabei [einen Mann], der Stuck an die Decke klebte. „Das Handwerk lebt“, rief Egon [dem Mann] zu. Der schaute ihn an und rief zurück: „Ja, aber vor allem braucht man [Augenmaß].“

Egon fragte sich, was das wohl alles kosten mag. Da muss der Baron ein hübsches Guthaben auf seinem Konto haben. „Kennen Sie Mona, unsere Chefredakteurin?“, fragte Egon den Baron. Dessen Blick verfinsterte sich noch mehr. „Ich kenne [niemanden mit] diesem Namen.“

A.1.3. Production experiment, text 3


Der Friedhof des Dorfes befindet sich etwas außerhalb. Von der Friedhofsmauer rankt Efeu.


Vor dem Haus steht eine Pappel, die ist 50 Jahre alt. Innen ist sie ganz morsch.
Als die Titanic unterging, entwickelte sich ein Sog, der alles in die Tiefe riss.
Die Höhle war schmaler als gedacht, sie mussten dicht am Boden krauchen, um vorwärts zu kommen.
Im nächsten Sommer machen wir auf der Insel Palau Urlaub, das ist ein kleiner Staat in Ozeanien.
Das Wetter ist heute sonnig und schön, ich würde so gern ein Eis essen gehen.
Der Junge stand auf der Klippe. Von dort blickte er über das Meer.
Der Zug fährt um 9.13 Uhr ab Allensbach und ist [fünfzehn Minuten] später in Konstanz.
Du willst wissen, wo die Pappe liegt? Die Pappe liegt im Mülleimer.
Selbstverständlich trank der Generalmajor Whisky, so wie es Männer von seinem Rang tun.
Wie oft habe ich dir gesagt, du sollst nicht diesen Dreck rauchen!
Der Mann fuhr schnell. Sie erreichten ruck zuck Lippe, einen Landkreis in Nordrhein-Westfalen.
Gestern habe ich im Kino einen alten Film mit Greta Garbo genossen. Der Freund seines Vaters hatte einen schweren Unfall gehabt. Seit zwei Tagen lag er im Koma.
Das Fußballspiel lief nun in der 91. Minute, doch der Ball wollte einfach nicht ins Tor gehen.
Monika und Jana waren einkaufen gewesen, jede in einem anderen Geschäft. Als sie später beim Kaffee zusammensaßen und sich ihre Einkäufe ansahen, merkten sie, dass sie den gleichen grünen Slip gekauft hatten.
Er wird die Zigarette nur halb rauchen, das macht er immer so - keine Ahnung, warum.
Unsere neue Kuh hat ihr erstes Kalb geboren. Wir werden es Mimi taufen.
Zuerst brauchen wir einen Plan. Sonst können wir den Bankraub vergessen.
Heute Abend wird gefeiert. Wir treffen uns in Rudis Kate.
Der Skalp eines Feindes stellte für Indianer eine wichtige Trophäe dar.
Die Tankstelle hatte gerade geöffnet. Peter schaute auf die Preistafel, während er vor seinen Tank trat. Über ein Euro der Liter Diesel! Er überlegte kurz, ob er voll oder nur halb voll tankt.
Die alte Frau hatte sich offenbar verlaufen. Peter rief ihr zu: "Da entlang, Oma!" Aber sie hörte nicht hin.
Ein Gewitter zog auf. Aus der Ferne war ein leises Grollen zu hören.
Peter hatte bald herausgefunden, warum die Maschine nicht lief: Die Klappen im
vorderen Teil [funktionierten nicht].
Die misshandelte Frau konnte [ihrem Mann nicht] mehr glauben, dass er sich ändern würde. Seine Versprechen waren jedes Mal nicht mehr als ein einziges Lippenbekenntnis gewesen.
Susi hatte zum Geburtstag neue Inline-Skater geschenkt bekommen. Jetzt wollte sie unbedingt am Hang rollen üben.
In Lappland sind die Männer besonders stark, heißt es. Wie stark Lappen wirklich sind, findet man spätestens beim Wettkampf im Schnee mit ihnen heraus.
Die drei Freunde hatten sich angewöhnt, ihren Grappa lauwarm zu trinken. So wurde man schneller beschwipst.
Die Gruppe wollte Skat spielen, doch es fehlte der dritte Spieler.
Seit sie ein eigenes Konto hatte, wurden ihr im Geschäft an der Fischerstraße zwei Prozent Skonto gewährt.
Mit Pfeil und Bogen war Gudrun unschlagbar. Sie war seit zehn Jahren Bogenschützin.
A.1.4. Production experiment, reading text for the second singleton-geminate analysis

Measured segments are shown as bold.

| Moritz liebte Sina.    | Micha wollte gehen. |
| Konrad schubste Anna.  | Linda rannte weiter. |
| Simon schätzte Rita.   | Oswin kannte Lola.  |
| Sarah warnte Anton.    | Rosi pflegte Martin. |
| Heidi kannte Wilhelm.  | Egon fragte Arnim.  |
| Wilhelm notierte Daten.| Arnim nieste hörbar. |
| Anton mochte Lisa.     | Martin mischte Karten. |
| Wilhelm mogelt öfter.  | Arnim merkte alles.  |
| Simon morgelt ständig. | Martin nickte eifrig. |

| Nadja trällert Lieder.  | Nele konnte sprechen. |
| Martha flüstert leise.  | Mechthild spielte Geige. |
| Andi schaut rauber.     | Henri pfückte Blumen. |
| Irmingard sprintet schneller. | Irmintraut backte Kuchen. |
| Tarzan brüllte lauter.  | Tristan übte fleißig. |
| Adam zwiefelt weiter.   | Anselm turnte sicher. |
| Berta lobte Hermann.    | Siegfried mochte Maren. |
| Gerhard hörte Adam.     | Heiko kannte Anselm. |
| Adam nickt manchmal.    | Friedhelm nickte rüber. |
| Hermann malt Bilder.    | Maren merkte alles. |
| Adam machte Fratzen.    | Anselm meldet Feuer. |
| Hermann nähte Hosen.    | Tristan nähte Kleider. |
A. Appendix

A.1.5. Perception experiment, additional recording

As before, targets are underlined. Targets that have been used for the study of silent interval perception, have a hash separating those phonemes, where the silent interval was spliced into.

1. Am liebsten esse ich Hering, Hecht, Aal und Rot#barsch.
2. Der Skalp eines Feindes stellt eine besondere Trophäe dar.
4. Der Aufbau von Windkraftanlagen wird jetzt nicht mehr so stark subventioniert wie im letzten Jahr.
5. Anna hatte sich beim Fuß#ball das Knie verletzt.
6. Anna wollte Süßes kandieren, um die Kinder zu überraschen.
7. Beim Kegeln kommt es darauf an die Kugel knapp innen zu führen.
8. Benzin wird in diesen Tagen rasch teuer.
9. Bis Kassel sind es noch zweihundert Kilometer.
11. Flatternde Tücher machen Rinder wahn#sinnig, nicht die rote Farbe.
12. Das Haus, das ihnen der Makler zeigte, war zum Glück alt. Ein hochmodernes hätte ihnen nicht gefallen.
13. Das Kind rollt mit seinem Dreirad über die Straße.
14. Das Land hatte schon andere Krisen als diese verkraftet.
15. Roh trat er zu, als das Opfer am Boden lag.
16. Das Geschäft an der Fischerstraße gibt zwei Prozent Skonto.
17. Die Wachtel hatte vier Eier in ihrem Nest.
18. Denk niemals, dass ich so etwas tun würde.
20. Der Knoten war so fest, dass sie ihn nicht lösen konnte.
21. Der Lagebericht war auf Englisch verfasst.
22. Der Entführer hatte sein Opfer mit einem Knebel gefesselt.
23. Zum freiwilligen Aufräumen meldete sich keiner.
24. Der Mann sah Tom auf der anderen Straßenseite.
25. Wenn Du Dir ein neues Auto kaufst: Unbedingt auf Bau und Leistungsfähigkeit achten!
26. Wir treffen uns in Rudis Kat.
27. Der Zoo kam in Verruf, weil er seine Tiere nicht artgerecht hielt.
29. Die Schüler kehren ihren Klassenraum.
30. Die bis dahin stumme Lehrerin sprach nun doch.
31. Die Demonstranten stehen auf der Straße und skandieren immer dieselbe Parole.
32. Die Kasse leuchtet im Dunkeln.
34. Der Dieb erklomm rasch Turm und Dach, sprang auf das nächste Gebäude und entkam.
35. Am Morgen lag Nebel über der Stadt.
36. Die Lehrerin freute sich während der Klassenfahrt über das fröhliche Jugendleben.
37. Die Mutter hatte zwei Stunden gewartet, jetzt war die Suppe kalt.
38. Die Nacht war rabenschwarz.
39. Er wachte langsam auf, denn die Sonne blinzelte in sein Gesicht.
40. Rot ist die Farbe der Leidenschaft.
41. Die Pinnen sind lange Hebel, mit denen bei Segelbooten das Ruder bedient wird.
42. Einen Troll gibt es nur im Märchen.
43. Englische Literaturverfilmungen sind in der Regel werkgetreu.
44. Susi eröffnet in der Bank am Markt ein neues Konto.
45. Er hat den Stein mit ganzer Kraft losgetreten.
47. Nach dem Einkaufen bemerkten Ines und Jana, dass sie den gleichen grünen Slip gekauft hatten.
48. Der Rinderwahnsinn entpuppte sich als hoch ansteckende Krankheit.
49. Die Pappel vor dem Haus ist schon fünfzig Jahre alt.
50. Der Junge beobachtet Raben im Garten.
51. Man sollte darauf achtgeben, wie der Strom produziert wird.
52. Er sah verwundert, dass auf dem Tisch Tiere saßen.
53. Die regionale Wasserwirtschaft schreibt schwarze Zahlen.
54. Dass in der Feuerwehr Männer arbeiten, weiß jeder. Feuerwehrfrauen gibt es aber auch.
55. Unsere Kuh hat ein neues Kalb geboren.
56. Als Störenfried sieht er sich nicht.
57. Die Torte war aus Eis und Schokolade.
58. Die Seefahrer um Kolumbus waren vom Entdeckergeist angetrieben.
59. Es ist erstaunlich, wie stark Riesen im Märchen sind.
60. Ich trage am liebsten Bluse und Rock.
61. Die neue Kollegin heißt Mona Tölle.
62. Das Mädchen merkte bald, dass die Treueschwüre ihres Freundes nicht mehr als ein bloßes Lippenbekenntnis gewesen waren.
63. Der neue Schlossherr ist für sein ausschweifendes Leben bekannt.
64. Ihre Hände lösten sich aus der Verkrampfung.
65. Mit lockerem Traben beginnt das Laufttraining.
66. Ihre Wäsche klammern sie immer ganz fest an die Leine, wenn ein Sturm aufzieht.
67. In der Nacht sind alle Raben schwarz.
68. Das **Lot** ist ein Gerät zum Messen der Meerestiefe.

69. In Großbritannien wird **Händel** oft als Nationalkomponist gefeiert, obwohl er eigentlich Deutscher war.

70. In der **Regel** wird der Sommer in Italien sehr heiß.

71. In ihrer **Jugend** leben viele Menschen über ihre Verhältnisse.

72. In letzter Zeit wirkte sie irgendwie **kraftlos**.

73. In diesem **Monat** lief sein Geschäft nicht so gut.

74. Könnt Ihr euren **Kram** nicht woanders hinlegen?

75. Lena hatte erst vor drei Monaten den **Führerschein** gemacht, doch sie saß schon sicher am **Steuer**.

76. Während der **Holunderblüte** präsentiert sich die Region von ihrer besten **Seite**.

77. Ein fröhliches **Hallo** tönte durch den Saal, als der Ehrengast eintrat.

78. Sein **Bart** bekam langsam graue Stellen.

79. Dass Wind und **Wasser** wirtschaftlich sind, wird von Befürwortern der Kernkraftwerke verneint.

80. Die Kinder übten **Latein**. [two targets: rege_L and Latein]

81. Meistens verkehrt der Bus **fahrplanmäßig**.

82. Mit dem **Regelwerk** kannte er sich aus.

83. Monika bat ihn barsch, Pannen zu vermeiden.

84. Sein Buch mit dem schönen Titel “Musst Du wieder stören, Friedhelm?” schlägt alle Verkaufsrekorde.

85. Nach dem vielen Regen ist die **Saat** sehr gut aufgegangen.

86. Die **Feuerwehrmänner** waren innerhalb von sieben Minuten an der Unfallstelle.

87. Die Statuen zerstörten die **Entdecker** geistlos.

88. Peter war erstaunt, wie dick **Rahm** wird, wenn man ihn lange genug schlägt.

89. Wie alt **Orte** sind, kann man in der Chronik im Gemeindehaus nachlesen.
90. Va#ter meint, der Ast rage zu weit in den anderen Garten.
91. Von der Lage berichtet unsere Korrespondentin in Russland.
92. In dieser Bar tummeln sich viele berühmte Gäste.
93. Von seiner Zigarette blieb nur ein Stummel übrig.
94. Auf seinem Schloss herrscht er wie ein Des#pot.
95. Wäscheklammern aus Holz färben manchmal auf die Kleidung ab.
96. Wegen der Bauarbeiten änderte sich der Busfahrplan wöchentlich.
97. Weil es ihm in Berlin nicht mehr gefiel, zog er in eine andere Stadt.
98. Wenn die Stiere durch die Straßen von Pamplona jagen, sollte man lieber zuhause bleiben.
99. Wir spannen ein Zelt über die Wiese, damit wir auch bei schlechtem Wetter feiern können.
100. Marias Holunder blühte in diesem Jahr einen Monat eher.
101. Wie stark Noten desselben Verlegers voneinander abweichen, ist unglaublich.
A.2. List of Experimental Materials

A.2.1. List of Targets

A.2.1.1. Perception experiment 1

During reading, participants tended to devoice target words ending in η. If that happened, the uttered sequence is given in brackets. In addition to that, items with prosody cues (natural word and merger with different intonation contours) are indicated in the first column as “+ pros. cue”.

See following page.
<table>
<thead>
<tr>
<th>Word boundary marker</th>
<th>merged word</th>
<th>original sequence</th>
<th>English</th>
<th>natural word</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>single</td>
<td>möchte so gern</td>
<td>wie es Männern von seinem Rang tun</td>
<td>would like so much</td>
<td>Sog</td>
<td>suction</td>
</tr>
<tr>
<td>single</td>
<td>rankt</td>
<td>wie es Männern von seinem Rang tun</td>
<td>like men of his rank do it</td>
<td>rankt</td>
<td>to entwine, 3. p.sg.</td>
</tr>
<tr>
<td>single</td>
<td>rolen</td>
<td>am Hang rolen üben</td>
<td>to practice inline skating downhill</td>
<td>rolen</td>
<td>to tumble</td>
</tr>
<tr>
<td>single</td>
<td>rauchen</td>
<td>Dreck rauchen</td>
<td>to smoke crud</td>
<td>rauchen</td>
<td>to crouch</td>
</tr>
<tr>
<td>single</td>
<td>z.Kalb</td>
<td>erstes Kalb</td>
<td>first calf</td>
<td>Kalb</td>
<td>scalp</td>
</tr>
<tr>
<td>single</td>
<td>z.lip</td>
<td>ein einziges Lippenbekenntnis</td>
<td>just lip service</td>
<td>lip</td>
<td>slip</td>
</tr>
<tr>
<td>single</td>
<td>Kat</td>
<td>Rudl’s cabin</td>
<td>name of a place</td>
<td>Kat</td>
<td>name of a place</td>
</tr>
<tr>
<td>single</td>
<td>Konto</td>
<td>eigenes Konto</td>
<td>one’s own bank account</td>
<td>Konto</td>
<td>allowance</td>
</tr>
<tr>
<td>single</td>
<td>wasser wirtschaft</td>
<td>dass Wind und Wasser wirtschaftlich</td>
<td>that wind and water are profitable</td>
<td>wasser wirtschaft</td>
<td>water economics</td>
</tr>
<tr>
<td>single</td>
<td>achtgeben</td>
<td>davon kaum ich [...] genau acht geben</td>
<td>I can give exactly eight of that</td>
<td>achtgeben</td>
<td>to take care</td>
</tr>
<tr>
<td>single</td>
<td>stören</td>
<td>musst du wieder stören, Friedhelm</td>
<td>do you have to interfere again, Friedhelm</td>
<td>stören</td>
<td>troublemaker</td>
</tr>
<tr>
<td>single</td>
<td>Schlossherr</td>
<td>auf seinem Schloss herrschen</td>
<td>rules in his palace</td>
<td>Schlossherr</td>
<td>lord of the castle</td>
</tr>
<tr>
<td>single</td>
<td>Augenmaß</td>
<td>mit den Augen maß</td>
<td>measured by the eye</td>
<td>Augenmaß</td>
<td>sense of proportion</td>
</tr>
<tr>
<td>multi</td>
<td>rücken</td>
<td>rück endlich zur Seite</td>
<td>move aside now</td>
<td>Rücken</td>
<td>back (body)</td>
</tr>
<tr>
<td>multi</td>
<td>Reporter</td>
<td>die Chefin, deren Report er</td>
<td>the boss (fem.) whose report he</td>
<td>Reporter</td>
<td>reporter</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>mag er</td>
<td>Journalisten mag er schon gar nicht</td>
<td>journalists he does not like at all</td>
<td>mag</td>
<td>skinny</td>
</tr>
<tr>
<td>multi</td>
<td>Kodak</td>
<td>die Kodak Klingt</td>
<td>the coda (mus.) sounds</td>
<td>Kodak</td>
<td>name of a company</td>
</tr>
<tr>
<td>multi</td>
<td>Pappel</td>
<td>Pappel liegt</td>
<td>poplar tree</td>
<td>Pappel</td>
<td>poplar</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>lippe</td>
<td>das Handwerk lebt</td>
<td>craftsmanship is alive</td>
<td>lippe</td>
<td>to stick, 3rd p. sg.</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>lippe</td>
<td>sie erreichten ruck zuck Lippe</td>
<td>they quickly reached Lippe (name of a German region)</td>
<td>lippe</td>
<td>cliff</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>tanкт</td>
<td>von seinem Tank trat</td>
<td>stepped in front of his (car’s) tank</td>
<td>tanкт</td>
<td>to fuel, 3rd p. sg.</td>
</tr>
<tr>
<td>multi</td>
<td>Oma</td>
<td>da entlang, Oma [entla ng]</td>
<td>this way, granny</td>
<td>Oma</td>
<td>oma</td>
</tr>
<tr>
<td>multi</td>
<td>All</td>
<td>ab Allensbach</td>
<td>from Allensbach (small but well-known town in the South of Germany)</td>
<td>All</td>
<td>ball</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>Mona T</td>
<td>Mona Tölle</td>
<td>first name and family name</td>
<td>Mona</td>
<td>month</td>
</tr>
<tr>
<td>multi</td>
<td>Lappen</td>
<td>wie stark Lappen</td>
<td>how strong lapps</td>
<td>Lappen</td>
<td>lids</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>rauchen</td>
<td>nur halb rauchen</td>
<td>to smoke only half (of the cigarette)</td>
<td>rauchen</td>
<td>to need</td>
</tr>
<tr>
<td>multi</td>
<td>zu kommen</td>
<td>zu kommen</td>
<td>to come</td>
<td>kommen</td>
<td>train</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>Palau</td>
<td>Grappa lauwarm</td>
<td>(drink) Grappa lukewarm</td>
<td>Palau</td>
<td>name of an island</td>
</tr>
<tr>
<td>multi</td>
<td>Bogen</td>
<td>Film mit Greta Garbo genesen</td>
<td>enjoy a film with Greta Garbo</td>
<td>Bogen</td>
<td>bow</td>
</tr>
<tr>
<td>multi</td>
<td>der Betrieb von</td>
<td>die Betätigung</td>
<td>the operation of</td>
<td>der</td>
<td>rough</td>
</tr>
<tr>
<td>multi</td>
<td>die Belüftung</td>
<td>die Belüftung</td>
<td>the ventilation</td>
<td>die</td>
<td>thieves</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>auf Bau</td>
<td>unbedingt auf Bau und [...] achten</td>
<td>be sure to pay attention to the build-up and</td>
<td>auf Bau</td>
<td>construction</td>
</tr>
<tr>
<td>multi (+ pros. cue)</td>
<td>gut haben</td>
<td>Sie müssen es gut haben</td>
<td>you must be well-off</td>
<td>Gut haben</td>
<td>credit balance</td>
</tr>
</tbody>
</table>
A.2.1.2. Perception experiment 2

Items with silent intervals carry a hash which shows the location of insertion. See following page.
<table>
<thead>
<tr>
<th>Wb-marker</th>
<th>merged word</th>
<th>original sequence</th>
<th>English</th>
<th>natural word</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>glottal stop</td>
<td>Sekt trank einer</td>
<td>one drank sparkling wine</td>
<td>kein</td>
<td>no one</td>
<td></td>
</tr>
<tr>
<td>glottal stop</td>
<td>zum Glück alt</td>
<td>thankfully old</td>
<td>kalt</td>
<td>cold</td>
<td></td>
</tr>
<tr>
<td>glottal stop</td>
<td>dank ehrenamtlicher Helfer</td>
<td>thanks to volunteers</td>
<td>kehren</td>
<td>sweep</td>
<td></td>
</tr>
<tr>
<td>glottal stop</td>
<td>wie alt Orte sind...</td>
<td>how old places are</td>
<td>Torte</td>
<td>cake</td>
<td></td>
</tr>
<tr>
<td>glottal stop</td>
<td>...die Kugel knapp innen zu führen</td>
<td>to lead the ball close towards the inner line</td>
<td>Pinnen</td>
<td>tillers</td>
<td></td>
</tr>
<tr>
<td>/l/ wrong pos</td>
<td>ein bloßes Lippenbekenntnis</td>
<td>just lip service</td>
<td>Slip</td>
<td>slip</td>
<td></td>
</tr>
<tr>
<td>/l/ wrong pos</td>
<td>die Pappe liegt</td>
<td>cardboard lies</td>
<td>Pappel</td>
<td>poplar</td>
<td></td>
</tr>
<tr>
<td>/l/ wrong pos</td>
<td>die Hände lagen</td>
<td>the hands lay</td>
<td>Händel</td>
<td>name of a composer</td>
<td></td>
</tr>
<tr>
<td>/l/ wrong pos</td>
<td>die stumme Lehrerin</td>
<td>the mute teacher</td>
<td>Stummel</td>
<td>stump</td>
<td></td>
</tr>
<tr>
<td>/l/ wrong pos</td>
<td>wachte langsam auf</td>
<td>woke up slowly</td>
<td>Wachtel</td>
<td>quails</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph fin</td>
<td>üben rege Latein</td>
<td>practice Latin eagerly</td>
<td>Regel</td>
<td>rule</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph fin</td>
<td>in dieser Bar tummeln</td>
<td>in this bar cavort</td>
<td>Bart</td>
<td>beard</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph fin</td>
<td>ein fröhliches Hallo tonte</td>
<td>a cheerful ‘hello’ sounded</td>
<td>Lot</td>
<td>perpendicular</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph fin</td>
<td>der Zoo kam in Verruf</td>
<td>the zoo’s reputation was tarnished</td>
<td>zog</td>
<td>moved, 3rd p. sg.</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph fin</td>
<td>er sah Tom</td>
<td>he saw Tom</td>
<td>Saat</td>
<td>seed</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph fin</td>
<td>roh trat er zu</td>
<td>he kicked brutally</td>
<td>rot</td>
<td>red</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>wie dick Rahm werden kann</td>
<td>how thick cream may get</td>
<td>Kram</td>
<td>stuff</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>das Kind rollt</td>
<td>the child rolls</td>
<td>Troll</td>
<td>troll</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>wie stark Noten [...] abweichen</td>
<td>how greatly musical notes differ</td>
<td>Knoten</td>
<td>knot</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>lag Nebel über der Stadt</td>
<td>fog lay over the city</td>
<td>Knebel</td>
<td>gag</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>denk niemals</td>
<td>never think</td>
<td>Knie</td>
<td>knee</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>wie stark Riesen im Märchen sind</td>
<td>how strong giants are in fairy tales</td>
<td>Krisen</td>
<td>crises</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>der Junge beobachtet Raben</td>
<td>the boy is watching ravens</td>
<td>traben</td>
<td>trot, v</td>
<td></td>
</tr>
<tr>
<td>fin stop aloph ini</td>
<td>Vater meint, der Ast rage</td>
<td>father thinks, the branch loomed</td>
<td>trage</td>
<td>carry, 3rd p. sg.</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>ein neues Kalb</td>
<td>a new calf</td>
<td>Skalp</td>
<td>scalp</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>ein neues Konto</td>
<td>a new bank account</td>
<td>Skonto</td>
<td>allowance</td>
<td></td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>treffen uns in Rudis Kate</td>
<td>Rudis’s cabin</td>
<td>Skat</td>
<td>name of card game</td>
<td></td>
</tr>
<tr>
<td>Wb-marker</td>
<td>merged word</td>
<td>original sequence</td>
<td>English</td>
<td>natural word</td>
<td>English</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------------</td>
<td>---------</td>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>skandieren</td>
<td>Süßes kandieren</td>
<td>to candy sweets</td>
<td>skandieren</td>
<td>to chant</td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>schüppen</td>
<td>auf dem Tisch Tiere</td>
<td>animals on the table</td>
<td>Stiere</td>
<td>bulls</td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>schüppen</td>
<td>bat ihm barsch, Pannen zu vermeiden</td>
<td>asked him harshly to avoid malfunctions</td>
<td>spannen</td>
<td>span,v</td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>schüppen</td>
<td>er erklomm rasch Turm</td>
<td>he quickly climbed up the tower</td>
<td>Sturm</td>
<td>storm, n</td>
</tr>
<tr>
<td>ini stop aloph med</td>
<td>schüppen</td>
<td>wird rasch teuer</td>
<td>gets expensive quickly</td>
<td>Steuer</td>
<td>tax</td>
</tr>
<tr>
<td>manipul stress</td>
<td>Busfahrplan</td>
<td>verkehrt der Bus fahrplannmäßig</td>
<td>the bus goes according to schedule</td>
<td>Busfahrplan</td>
<td>bus time schedule</td>
</tr>
<tr>
<td>manipul stress</td>
<td>Feuerwehr-Männer</td>
<td>das in der Feuerwehr Männer arbeiten</td>
<td>that men work in the fire brigade</td>
<td>Feuerwehrmänner</td>
<td>firemen</td>
</tr>
<tr>
<td>manipul stress</td>
<td>kraftlos</td>
<td>trat ihn mit aller Kraft los</td>
<td>kicked it loose with all his power</td>
<td>kraftlos</td>
<td>powerless</td>
</tr>
<tr>
<td>manipul stress</td>
<td>Jugendleben</td>
<td>in der Jugend leben</td>
<td>during their youth</td>
<td>Jugendleben</td>
<td>youth life</td>
</tr>
<tr>
<td>manipul stress</td>
<td>Lagebericht</td>
<td>von der Lage berichtet</td>
<td>reports about the situation</td>
<td>Lagebericht</td>
<td>position report</td>
</tr>
<tr>
<td>manipul stress</td>
<td>Wasserwirtschaft</td>
<td>dass Wind und Wasser wirtschaftlich</td>
<td>that wind and water are profitable</td>
<td>Wasserwirtschaft</td>
<td>water economics</td>
</tr>
<tr>
<td>manipul stress</td>
<td>Aufbau</td>
<td>auf Bau und [...] achten</td>
<td>be sure to pay attention to the build-up and</td>
<td>Aufbau</td>
<td>construction</td>
</tr>
<tr>
<td>manipul stress</td>
<td>Wäscheklammern</td>
<td>ihre Wäsche klammern sie immer</td>
<td>they always peg their laudry</td>
<td>Wäscheklammern</td>
<td>pegs</td>
</tr>
<tr>
<td>no wb marker</td>
<td>Schlossherr</td>
<td>auf seinem Schloss herrscht</td>
<td>rules in his palace</td>
<td>Schlossherr</td>
<td>lord of the castle</td>
</tr>
<tr>
<td>no wb marker</td>
<td>Holunderblüte</td>
<td>Marias Holunder blühte</td>
<td>Maria’s elderblossom bloomed</td>
<td>Holunderblüte</td>
<td>elder bloom</td>
</tr>
<tr>
<td>no wb marker</td>
<td>achtgeben</td>
<td>davon kann ich [...] genau achtgeben</td>
<td>I can give exactly eight of that</td>
<td>achtgeben</td>
<td>to take care</td>
</tr>
<tr>
<td>no wb marker</td>
<td>Störenfried</td>
<td>musst du wieder stören, Friedhelm</td>
<td>do you have to interfere again, Friedhelm</td>
<td>Störenfried</td>
<td>troublemaker</td>
</tr>
<tr>
<td>no wb marker</td>
<td>Entdeckergeist</td>
<td>zerstören die Entdecker geistlos</td>
<td>the discoverers destroyed heedlessly</td>
<td>Entdeckergeist</td>
<td>exploratory spirit</td>
</tr>
<tr>
<td>no wb marker</td>
<td>Regelwerk</td>
<td>in der Regel werkgetreu</td>
<td>usually true to the original</td>
<td>Regelwerk</td>
<td>rulebook</td>
</tr>
<tr>
<td>no wb marker</td>
<td>Rabenwahn</td>
<td>in der Nacht sind alle Raben schwarz</td>
<td>during the night all ravens are black</td>
<td>Rabenwahn</td>
<td>mad cow disease</td>
</tr>
<tr>
<td>no wb marker</td>
<td>Rinderwahn</td>
<td>Tücher machen Rinder wahnsinnig</td>
<td>cloths drive cows mad</td>
<td>Rinderwahn</td>
<td>mad cow disease</td>
</tr>
<tr>
<td>silent interval</td>
<td>Sei#te</td>
<td></td>
<td>page</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent interval</td>
<td>La#tein</td>
<td></td>
<td>Latin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent interval</td>
<td>erwider</td>
<td></td>
<td>realised, 3rd p. sg.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent interval</td>
<td>Va#ter</td>
<td></td>
<td>father</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent interval</td>
<td>Rot#barsch</td>
<td></td>
<td>redfish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent interval</td>
<td>De#spot</td>
<td></td>
<td>despot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent interval</td>
<td>Fuß#ball</td>
<td></td>
<td>football</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent interval</td>
<td>er#klomm</td>
<td></td>
<td>created, 3rd p. sg.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A. Appendix
A.2.2. Singleton-Geminate Targets

A.2.2.1. Test 1 (uncontrolled)

All nasal segments per item that were measured are shown as bold.

<table>
<thead>
<tr>
<th>Target</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>einem Monat</td>
<td>one month, dat.</td>
</tr>
<tr>
<td>mager</td>
<td>skinny</td>
</tr>
<tr>
<td>dem Mann</td>
<td>the man, dat.</td>
</tr>
<tr>
<td>Guthaben</td>
<td>credit balance</td>
</tr>
<tr>
<td>niemanden mit</td>
<td>no-one with</td>
</tr>
<tr>
<td>Augenmaß</td>
<td>sense of proportion</td>
</tr>
<tr>
<td>[...]Interview mit Baron von Rücken machen</td>
<td>to do an interview with Baron von Rücken</td>
</tr>
<tr>
<td>einen Mann</td>
<td>a man, acc.</td>
</tr>
<tr>
<td>Debatten mit</td>
<td>discussions with</td>
</tr>
<tr>
<td>fünfzehn Minuten</td>
<td>fifteen minutes</td>
</tr>
<tr>
<td>funktionierten nicht</td>
<td>they did not function</td>
</tr>
<tr>
<td>[konnte] ihrem Mann nicht [mehr vertrauen]</td>
<td>could not trust her husband anymore</td>
</tr>
<tr>
<td>im Müllimer</td>
<td>in the bin</td>
</tr>
<tr>
<td>nun nicht</td>
<td>not now</td>
</tr>
<tr>
<td>schon nach</td>
<td>already after</td>
</tr>
</tbody>
</table>
A. Appendix

A.2.2.2. Test 2 (controlled)

Nasal segments that were measured for the main test (section 3.4.6), are shown as bold. A side study with medial singletons was carried out that examined medial singleton durations depending on syllable position (figure 3.9). Medial nasals in syllable-final position from the main test (bold) were compared with medial nasals in syllable-initial position (underlined).

<table>
<thead>
<tr>
<th>initial nasal singleton</th>
<th>medial nasal singleton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norbert</td>
<td>Konrad</td>
</tr>
<tr>
<td>Moritz</td>
<td>Samson</td>
</tr>
<tr>
<td>Nadja</td>
<td>Andi</td>
</tr>
<tr>
<td>Martha</td>
<td>Irmgard</td>
</tr>
<tr>
<td>Nicki</td>
<td>Linda</td>
</tr>
<tr>
<td>Micha</td>
<td>Samson</td>
</tr>
<tr>
<td>Nene</td>
<td>Henri</td>
</tr>
<tr>
<td>Mechthild</td>
<td>Irmtraud</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>final nasal singleton, no IP bound</th>
<th>final nasal singleton at IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolfram</td>
<td>Anton</td>
</tr>
<tr>
<td>Simon</td>
<td>Wilhelm</td>
</tr>
<tr>
<td>Tarzan</td>
<td>Hermann</td>
</tr>
<tr>
<td>Adam</td>
<td>Adam</td>
</tr>
<tr>
<td>Osew</td>
<td>Martin</td>
</tr>
<tr>
<td>Argim</td>
<td>Arnim</td>
</tr>
<tr>
<td>Tristan</td>
<td>Marien</td>
</tr>
<tr>
<td>Anselim</td>
<td>Anselim</td>
</tr>
</tbody>
</table>

- **contact geminates**
  - Wilhelm notiert (W notes down)
  - Anton mochte (A liked)
  - Wilhelm mogelt (W cheats)
  - Simon norgelt (S quarrels)
  - Adam muschelt (A mumbles)
  - Hermann malte (H painted)
  - Adam machte (A made)
  - Hermann nahlte (H sewed)
  - Argim niste (A sneezed)
  - Martin mischte (M mixed)
  - Argim merkte (A realised)
  - Martin nickte (M nodded)
  - Friedhelm nickte (N nodded)
  - Marien merkte (M realised)
  - Anselim meldet (A reports)
  - Tristan nahlte (T sewed)