Incomplete Neutralization –
a cognitive artifact of lexical co-activation?

Master Thesis

for attainment of the academic degree of

Master of Arts

presented by

Timo Benjamin Röttger

Supervisor: Prof. Dr. Martine Grice
# Content

**List of Figures**

<table>
<thead>
<tr>
<th>Chapter I</th>
<th>Introduction - Contrast and Incomplete Neutralization</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter II</td>
<td>Incomplete Neutralization of the final voicing contrast</td>
<td>6</td>
</tr>
<tr>
<td>2.1</td>
<td>Empirical evidence for Incomplete Neutralization</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>Incorporation of IN into traditional phonological accounts</td>
<td>8</td>
</tr>
<tr>
<td>2.3</td>
<td>Methodological concerns</td>
<td>11</td>
</tr>
<tr>
<td>2.3.1</td>
<td>General methodological shortcomings</td>
<td>11</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Hyperarticulation</td>
<td>12</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Orthography bias</td>
<td>13</td>
</tr>
<tr>
<td>2.4</td>
<td>The settling of the debate?</td>
<td>16</td>
</tr>
<tr>
<td>Chapter III</td>
<td>Towards an explanation</td>
<td>19</td>
</tr>
<tr>
<td>3.1</td>
<td>Exemplar-based models</td>
<td>19</td>
</tr>
<tr>
<td>3.2</td>
<td>Incomplete Neutralization as a cognitive by-product?</td>
<td>22</td>
</tr>
<tr>
<td>3.3</td>
<td>Towards an answer</td>
<td>25</td>
</tr>
<tr>
<td>Chapter IV</td>
<td>Empirical validation</td>
<td>28</td>
</tr>
<tr>
<td>4.1</td>
<td>Method</td>
<td>28</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Participants and experimental procedure</td>
<td>28</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Speech material</td>
<td>31</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Acoustic analyses of input stimuli</td>
<td>32</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Stimuli manipulation</td>
<td>32</td>
</tr>
<tr>
<td>4.1.5</td>
<td>Stimuli norming</td>
<td>33</td>
</tr>
<tr>
<td>4.1.6</td>
<td>Stimuli presentation</td>
<td>34</td>
</tr>
</tbody>
</table>
4.1.7 Acoustic analyses of production data 34
4.1.8 Statistics 35
4.2 Results and Discussion 36
  4.2.1 Omnibus Model (first production of all stimuli) 37
  4.2.2 Subset Model (repeated stimuli) 39
4.3 Summary 44

Chapter V Putting the picture together 46
  5.1 To be or not to be? The representation question 47
  5.2 Incomplete Neutralization in perception 48
  5.3 Incomplete Neutralization and lexical organization 50

Chapter VI Concluding remarks 54

References 58
Appendix A (test stimuli) 68
Appendix B (demonstration stimuli) 70
Appendix C (training stimuli) 71
Appendix D (R output omnibus model & subset model) 72

Erklärung 74
List of figures

Figure 1: Schematic waveforms of the word pair Bad/bat ‘bath/begged’. The onset of the vowel preceding the stop is represented through the white rectangle, the smaller gray rectangle is the voicing-into-the-closure duration the straight line is the closure duration, and the triangle represents the duration of the stop burst visible on the oscillogram. Figure reprinted from Port & O’Dell (1985: 459).

Figure 2: Vowel durations for underlying voiced (white) and voiceless (gray) stops in the neutralizing context for all 16 speakers; top row: speakers 1 to 8, bottom row, speakers 9 to 16. Figure reprinted from Röttger et al. (2011: 1724).

Figure 3: Acoustic waveform, spectrogram, f0 contour, and GToBI annotation of the auditory stimulus Peter weiß nun, wie die Bauge aussehen ‘Peter knows now what the Bauge looks like’.

Figure 4: Acoustic waveform, spectrogram, f0 contour, and GToBI annotation of the response sentence in the demonstration session denn nur der Men sieht so aus ‘as only the Men looks like this’.

Figure 5: Manipulation conditions.

Figure 6: Vowel durations for underlying voiceless (white; mean=123.72ms, SD=31.16ms) and voiced (gray; mean=127.86ms, SD=31.68ms) stops in the neutralizing context.

Figure 7: Vowel durations for underlying voiceless (white) and voiced (gray) stops in the neutralizing context for all 16 speakers; left: speakers 1 to 8; right: speakers 9 to 16.

Figure 8: Mean values and SE of proportional IN effects for all four conditions.

Figure 9: Proportional IN effects for (1) enhanced and (4) reversed condition; the dashed line represents no difference between underlying voiced and voiceless stops.

Figure 10: Mean proportional IN effects for enhanced (black bars) and reversed condition (gray bars) over occurrences. For comprehensibility, no variance is given here (SD over 14% for all conditions).

Figure 11: Effects in relation to each other (= proportional IN effect (enhanced condition) / proportional IN effect (reversed condition)). Bold line indicates 1:1 relation.
Chapter I - Introduction. Contrast and Incomplete Neutralization

Introduction
Contrast and Incomplete Neutralization

“The final question lies at the doorstep of phonemic theory:
What phonetic differences are contrastive in a given language?”
Goldsmith (1996: 1)

The concept of contrast is the central element in phonological theory. It played an important role in the major phonological schools of the twentieth century (de Saussure, 1916; Hocket, 1955; Sapir, 1933; Trubetzkoy, 1939) and it still does as is demonstrated by the statement above made by John Goldsmith in his introduction to The Handbook of Phonological Theory.

Two sounds are considered phonologically contrastive if they are in opposition with each other, i.e., if they are capable of distinguishing the meaning of two lexical forms in a particular language. Compare the minimal pairs given in (1):

(1i)  ENGLISH  
betting  /betɪŋ/  
bedding  /bedɪŋ/

(1ii) GERMAN  
En te  ‘end’  /ɛntə/  
En de  ‘duck’  /ɛnda/  

Both English and German contrast stops intervocally through voicing (among other features). Early theories of phonology assumed that out of the vast amount of phonetic
signals, only a small subset of phonetic properties are contrastive in a given language (Chomsky & Halle, 1968; Hocket, 1955; Kiparsky, 1982; Sapir, 1933; Trubetzkoy, 1939). Contrasts were represented in terms of feature differences between $+F_x$ and $-F_x$ for a finite set of features. In the cases above (cf. (1)), a feature [±VOICE] was assumed that contrasts the two lexical meanings from each other. When a phonological contrast is abrogated, neutralization obtains. In feature systems, neutralization corresponds to the elimination of the feature distinction.

A prominent example of neutralization is final devoicing: languages, such as Catalan, Dutch, Polish, Russian, Turkish, and German contrast voiced obstruents intervocally but neutralize the contrast in syllable-final or word-final positions in favor of voiceless segments (cf. (2)).

\[
\begin{align*}
(2i) & \quad \text{GERMAN} & \text{Tote} & \text{‘dead people’} & /tɔːtə/ \\
& & \text{Tode} & \text{‘deaths’} & /tɔːdə/ \\
(2ii) & \quad \text{GERMAN} & \text{tot} & \text{‘dead’} & /tɔːt/ \\
& & \text{Tod} & \text{‘death’} & /tɔt/ \\
\end{align*}
\]

This asymmetric distribution is described in terms of final devoicing\footnote{The German phonological contrast is also described in terms of lenis versus fortis (cf. Kohler, 1984). Hence, the process of final devoicing could be described as final fortitioning. To avoid confusion, I will use the terms voiced, voiceless and final devoicing throughout the paper, because this is the terminology that has been used in the Incomplete Neutralization debate.} of the voiced segment. German final devoicing has been called the “universally recognized archetype of phonological neutralization” (Fourakis & Iverson, 1984: 141), and it has been described as a “classic example of a phonological rule” (Wiese, 1996: 204). In terms of generative rules this neutralization has been formalized in the following way:

\[
(+\text{OBSTRUENT}) \rightarrow (-\text{VOICE}) / [-]_{\text{Coda}}
\]

This rule causes Tod (‘death’) and tot (‘dead’), that differ underlyingly\footnote{Again, to avoid confusion with the terminology used in the IN debate I will use the term “underlying” throughout this paper. I do not commit to any theoretically motivated idea of representation or derivation associated to this term. I simply refer to the voicing status of the corresponding intervocalic segment within the corresponding morphological paradigm.} in the voicing status of the final consonant, to become phonetically homophonous. Behind such a rule
lies the assumption that final devoicing – as a purely phonological phenomenon – is phonetically complete, with the two segments surfacing as absolutely indistinguishable from one another. However, numerous researchers have contested that many languages that traditionally were thought to exhibit complete neutralization actually exhibit *incomplete* neutralization (henceforth IN). Based on detailed acoustic studies, it has been argued that there are small acoustic and articulatory differences between words such as *Tod* and *tot* (for German: Charles-Luce, 1985; Dinnsen, 1985; Dinnsen & Garcia-Zamor, 1971; Mitleb, 1981; O’Dell & Port, 1983; Port & Crawford, 1989; Port, Mitleb & O’Dell, 1981; Port & O’Dell, 1985; Röttger, Winter & Grawunder, 2011). In particular, it has been claimed that the devoiced stop is actually “partially voiced” – in other words, the true contrast of voiced versus voiceless stops in intervocalic position becomes reduced to a “semicontrast” in final position, with only some weak phonetic cues remaining. It has further been shown that listeners are sensitive to IN, since they can tell apart “devoiced” stops from genuinely voiceless ones with above-chance accuracy (Kleber, John & Harrington, 2010; Port & Crawford, 1989; Port & O’Dell, 1985; Röttger, Winter & Grawunder, 2011).

Such IN findings are not only restricted to final devoicing. Acoustic studies have revealed incomplete neutralization for spirantization in Eastern Andalusian Spanish (Bishop, 2007; Gerfen, 2002; Gerfen & Hall, 2001), consonant deletion in Turkish (Dinnsen, 1985), liquid neutralization in Puerto Rican Spanish (Simonet, Rohen-Madrazo & Paz, 2008), and flapping of intervocalic alveolar stops in American English, which has been extensively investigated on the background of IN (for an overview, see de Jong, 2011). A similar phenomenon called “Near Mergers” refers to the situation that a speaker consistently reports that two categories of sounds are the same, but consistently differentiate them in production at a better than chance level. Near mergers

---

3 The difference between *Incomplete Neutralization* and *Near Mergers* is primarily one that has to do with the history of investigation of that contrast (Manaster Ramer, 1996). In the case of Incomplete Neutralization, the majority of researchers thought the contrast to be neutralized, but then production studies found acoustic differences. In the case of near mergers, no such prior assumption was made, and often there was a recent sound change that resulted in the near merger. There is no difference in the nature of the evidence that is taken to support IN, and the evidence that is taken to support near mergers. While Near Meger is commonly used in the context of diachronic sound change, Incomplete Neutralization refers to a synchronic observation. Recently, Yu (2007, 2011) discussed these concepts within the same account.
have been reported for vowels in many English dialects (Di Paolo, 1988; Harris, 1985; Labov, 1971; Nunberg, 1980; Trudgill, 1974). E.g. Labov et al. (1972: ch. 6) have demonstrated that New York speakers differentiate source and sauce acoustically in production but report no distinction between them in perception. Near mergers are not only restricted to segmental contrasts. Yu (2007) reported on tonal near mergers in Cantonese. He has demonstrated that derived mid-rising tones show small but statistically significant differences in $f_0$ from underived mid-rising tones. All of the above cited studies provide evidence that the loss of a contrast can be incomplete. Some researchers even went as far as to propose that “many putative neutralizations, when examined more carefully, may be shown to be non-neutralizing on all accounts (...)” (Dinnsen, 1985: 277), or stating that researchers should consider the “radical hypothesis” that true neutralization never occurs (Port & O’Dell, 1985: 466). At least it can be said that it is a fairly common phenomenon that contrasts that are putatively lost exhibit a small but systematic difference in production.

Accepting IN as a robust phenomenon would mean that, in the case of IN of final devoicing, final obstruents are realized with “partial voicing” as a kind of in-between contrast, a “semicontrast.” Most of the early proponents of IN assumed that the facts about this phenomenon need to be integrated into a generative model of phonology: they addressed IN findings by adding new features to the underlying structure or adding new rules to the derivation process. These approaches probably, to many phonologists, would appear post-hoc explanations that do not serve any predictive purpose and that, moreover, are undesirable because they would lead to a proliferation of features or rules (e.g. Charles-Luce, 1985; Port & Crawford, 1989; Port & O’Dell, 1985).

While these approaches are based on a view of lexical organisation and access that assumes little redundancy in the lexicon, there are alternative accounts that assume a rich lexicon full of interconnected information (e.g. Bybee, 2001; Pierrehumbert, 2001). This view on lexical organisation offers promising explanations for IN phenomena, which, moreover, make testable predictions. This is, striktly speaking, a conception of the lexicon that could allow researchers to formulate an account based on lexical co-activation of morphological neighbors influencing the production of corresponding neutralized forms.
Regardless of the right explanation, it seems clear that IN lies at the heart of our understanding of the phonetics-phonology interface: gradual, variant production differences which are difficult to perceive at all seem to correspond to discrete invariant phonological categories. The phenomenon of IN challenges fundamental assumptions about the phonological system and raises question about the true nature of contrasts and the loss of these. A clearer understanding of these phenomena and their cause will deepen our current knowledge about speech production and perception.

The work in hand focuses on IN of final devoicing in German. It makes following contribution to the debate: in Chapter II, I review the empirical evidence for incomplete neutralization; proposed phonological explanations are discussed critically and the reception of the scientific community is reviewed with a focus on methodological criticisms. In Chapter III, exemplar-based models of lexical organisation and lexical access are introduced. Within this model, the lexical co-activation explanation for IN is discussed. Moreover, in light of its empirical predictions, testable hypotheses are generated. In Chapter IV, I report on an experiment that tests these hypotheses. In Chapter V, I draw potential implications of the results and discuss a) the representational status of IN, b) the role of IN in perception, and c) the connection between language specific traits of the lexicon and IN. Finally, Chapter IV summarizes the discussion.
Chapter II

Incomplete Neutralization of the final voicing contrast

“German final devoicing is the universally recognized archetype of phonological neutralization [...]”
Fourakis and Iverson (1984: 141)

“German apparently does not have an abstract phonological rule of neutralization, despite almost a hundred years of assertions by linguists and German pedagogists that it does”
Port and Crawford (1989: 280)

In the remainder of this chapter I will discuss the empirical basis of IN (§2.1) and proposed accounts to incorporate IN into traditional phonological models (§2.2). It will turn out that traditional accounts addressing IN are fraught with problems. One response to that incompatibility is the rejection of IN as a genuine phenomenon: many scholars have tried to explain away the experimental results by pointing out apparent methodological errors. Section 2.3 will discuss these methodological concerns with a focus on two aspects in particular: hyperarticulation (§2.3.2) and the orthography bias (§2.3.3). Afterwards, I will discuss recent experimental results that suggest that IN is not solely due to methodological artifacts (§2.4).
2.1 Empirical evidence for Incomplete Neutralization

Production studies that have been taken to suggest the presence of IN have asked participants to read out minimal pairs like Tod (‘death’) and tot (‘dead’). For German, acoustic analyses have revealed small but significant differences between underlying voiced and voiceless obstruents (e.g. Charles-Luce, 1985; O’Dell & Port, 1983; Port, Mitleb & O’Dell, 1981; Port & O’Dell, 1985), the most important cues being the duration of the preceding vowel, the closure duration, the duration of the “voicing-into-closure”, as well as the burst duration and aspiration duration (cf. Fig. 1). Among these, the duration of the preceding vowel stands out to be the most reliable IN cue both across studies and across languages. Importantly, in German these cues are known to distinguish voiceless from voiced obstruents in intervocalic contexts as well (e.g. Keating, 1984; Kohler, 1984), and the direction of the differences resembles the non-neutralizing contrast. For example, vowels are shorter before underlying final voiceless stops than before final “devoiced” stops; however, the differences are generally much smaller than in intervocalic or initial positions.

![Figure 1: Schematic waveforms of the word pair Bad/bat ‘bath/begged’. The onset of the vowel preceding the stop is represented through the white rectangle, the smaller gray rectangle is the voicing-into-the-closure duration, the straight line is the closure duration, and the triangle represents the duration of the stop burst visible on the oscillogram. Figure reprinted from Port & O'Dell (1985: 459).](image)

Studies on other languages with final devoicing replicated IN effects for Dutch (Ernestus & Baayen, 2006; Warner et al., 2004), Catalan (Charles-Luce, 1993; Charles-Luce & Dinnsen, 1987; Dinnsen & Charles-Luce, 1984), Polish (Slowiaczek & Dinnsen, 1985; Slowiaczek & Szymanska, 1989) and Russian (Chen, 1970; Dmitrieva, Jongman & Sereno, 2010; Kanibolotskaia, 2008; Pye, 1986). The magnitude of the IN effect seems on the one hand dialect- and speaker-dependent (Piroth & Janker, 2004).
and on the other hand highly sensitive to phonetic, semantic, and pragmatic context (Charles-Luce, 1985, 1993; Ernestus & Baayen, 2006; Port & Crawford, 1989; Slowiaczek & Dinnsen, 1985).

It has been further shown that listeners can tell apart underlying voiced and voiceless obstruents above chance level even in the neutralized position (cf. Port & Crawford, 1989; Port & O’Dell, 1985; Slowiaiscek & Szymanska, 1989). However, the identification accuracies reported in these experiments are generally very low (e.g. 54% accuracy in Röttger et al. (2011); 59% in Port & Crawford (1989)) in comparison to the perception of contrasts in non-neutralized environments.

2.2 Incorporation of IN into traditional phonological accounts

Traditional theories assume that phonological representations in the lexicon are categorical while the implementation component computes highly gradient and variable outputs. So why does an output $X^{\text{minus}}$, derived from an underlying representation $X^{\text{plus}}$, display different surface realizations from an output $X^{\text{minus}}$ derived from an underlying representation $X^{\text{minus}}$? These models assume the flow of information to be unidirectional. In other words the phonetic implementation mechanisms cannot look back to the phonological representation. Thus, the information that $X^{\text{minus}}$ is derived from $X^{\text{plus}}$ is not available to the articulatory devices.

The phenomenon of IN poses serious difficulties for such accounts. Some authors, nevertheless, tried to incorporate IN into such models. Port and O’Dell (1985: 466) consider the possibility that a feature [VOICE-F] applies especially to the case of IN. The feature is phonetically implemented in a way that results in between voicing and the absence of voicing. Rather than adding a new feature, Charles-Luce (1985: 319ff.) considers the possibility that final devoicing is not a feature-changing rule [+OBSTRUENT] $\rightarrow$ [-VOICE] / _], but a feature-deletion rule. The resulting segment would be unspecified for [VOICE] and would be marked for having undergone the final devoicing process. This is called by Charles-Luce (1985) an ‘allophonic rule’ because the result is an allophone of the voiced stop that is somewhat intermediate in terms of voicing. Under this account, the final devoicing pattern would not be a neutralization phenomenon anymore.
Most rule-based views to address IN have tried to accommodate experimental results with the help of phonetic implementation rules that work on phonological input and produce phonetic “surface” output. Under this account, proposed by Port and O’Dell (1985), Germans are using an implementation rule that devoices at the syllable level. The phonological [+VOICE] feature that is assumed to be in the lexicon is left unchanged, but phonetically, only on the syllable level, a [-VOICE] gesture is realized. The realization of this [-VOICE] gesture is thought to depend on the context of utterance formation, e.g. whether clear speech is called for in a particular situation.

Another approach would be to alter the order of implementation rules (considered by Charles-Luce, 1985: 319), i.e. the phonological devoicing rule must be ordered after or simultaneously to the phonetic implementation rule. Note that this assumption is incompatible with the unidirectional hypothesis of the discussed feed-forward models. This would imply that the [VOICE] feature is still available to the implementation rule and is therefore allowed to affect the surface form. Dinnsen and Charles-Luce (1984: 57ff.) use the results of the IN studies to refute the idea that there is a separate phonological rule component ordered before the phonetic rule component altogether. They also consider the possibility of a phonological “shortening rule” that directly accounts for the main findings of almost all IN studies that vowels are shortened before underlying voiceless segments. This, however, does not account for other acoustic cues known to characterize IN, which in fact are the same as for domain medial contrasts between voiced and voiceless obstruents. Furthermore, this rule would be a phonological rule that already includes a lot of phonetic detail. The proposal is similar to Port and O’Dell’s (1985) implementation rule, but located at the phonological level.

In a more recent attempt to model IN in Optimality Theory, Oostendorp (2008) argues that output structures can be characterized in terms of Projection relations and Pronunciation relations. Projection relations are abstract relationships between segments and their features. Pronunciation relations are output relationships between the feature and the segment. They describe the actual phonetic output of a structure. This results in three different categories: there are segments that are underlingly voiced and pronounced voiced (e.g. [d] in [ʁæːdɐ] ‘wheels’), segments that are underlingly voiceless and pronounced voiceless (e.g. [t] in [ʁæːtɐ] ‘councils’ or [ɾaːt] ‘council’), and segments that are underlingly voiced but pronounced as voiceless (e.g. [ɾæːt] ‘council’).
[t] in [ʁa:t] ‘wheel’). Because of their structural differences, [ʁa:t] ‘council’ and [ʁa:t] ‘wheel’ will phonetically surface as different from each other. Note that this is very similar to Charles-Luce’s account of an allophonic rule discussed above.

To most phonologists, all of these attempts would probably be perceived as merely post-hoc explanations that do not serve any predictive purpose, and that moreover are undesirable because they would lead to a proliferation of features or rules. What is even more problematic is that these would have to be implemented in vastly different ways depending on context (Charles-Luce, 1985, 1993; Ernestus & Baayan, 2006; Port & Crawford, 1989; Slowiaczek & Dinnsen, 1985), and would vary a lot across speakers (e.g. Port & Crawford, 1989; Port & O’Dell, 1985; Warner et al., 2004). These are properties that are not considered desirable in feature- or rule based systems. From a cross-linguistic point of view, these accounts either predict all neutralizing processes in other languages to be incomplete, or they need to assume different neutralizing mechanisms for different languages, in which case one has to predict when the different processes apply. To sum up, it appears that traditional accounts to address IN are fraught with problems. Port and Crawford have recognized this and have claimed that findings of IN “pose a threat to phonological theory” (1989: 257; see also Port & Leary, 2005).

These problems have been responded to in several ways. Probably the predominant response was rejection of its relevance to linguistic theory, with a number of scholars pointing out apparent methodological errors (Fourakis & Iverson, 1984; Kohler, 2007; Manaster Ramer, 1996; Wiese, 1996) and trying to explain away the results. As pointed out by Fuchs (2005: 25), the controversy surrounding the notion of Incomplete Neutralization “became more and more a debate about methodological problems”. To gain a full understanding of the possible sources of IN, I will now discuss this methodological issues in detail.
2.3 Methodological concerns

2.3.1 General methodological shortcomings

A problematic aspect of previous studies that is in part discussed in the literature (Fourakis & Iverson, 1984; Manaster Ramer, 1996), is that many of the studies are based on astonishingly small sample sizes of speakers and/or items, e.g. both Charles-Luce (1985) and Port and Crawford (1989) had five participants, Fourakis and Iverson (1984) only had four. Charles-Luce (1985) tested only two stop minimal pairs and two fricative minimal pairs while Port and Crawford (1989) only had two pairs. Thus, on the one hand, studies that were not able to detect significant effects (hence reporting a null result) may have failed to show IN just because of too low statistical power. For Dutch, Warner et al. (2004) found small but significant differences in vowel and burst duration, whereas previous studies that used a smaller number of speakers and items did not (Baumann, 1995; Jongman et al., 1992). Thus, previous studies used sample sizes that did not demonstrate a “sufficiently good effort to disprove the null” (see Frick, 1995 for the Effort Criterion). With respect to IN, this means that any study attempting to “disprove” IN needs to have at least as many participants as the proponents of IN used in their study – and this was not the case with Fourakis and Iverson (1984). Additionally, one should recognize that most reported null results showed the predicted IN effects – albeit not significantly – in terms of numerical tendencies. Moreover, the studies that did find IN effects sometimes committed pseudoreplication, a statistical error that artificially inflates sample sizes and leads to an increase in Type I error rate, making it easier to find statistically significant results (Hurlbert, 1984; see Winter, 2011 for pseudoreplication in phonetic research). The term refers to the treatment of dependent data points as independent ones, for example the treatment of multiple samples from an individual as unrelated and uncorrelated samples. This error can lead to a drastic increase in Type I error rate, in some cases with up to a 50% probability of finding a spurious result that is indicated as being significant but that is actually due to chance.

Second, some of the prior studies used very rare or low frequency items. This is a problem, since low frequency items are often hyperarticulated (Whalen, 1991, 1992). Third, some studies of IN were conducted in America with German speakers who had
resided in the U.S. for quite some time. Given that English does not neutralize the final voicing contrast, these German speakers might have been influenced by their surrounding linguistic environment. This is particularly likely because participants were greeted and given instructions in English, and we know that the language and dialect of the experimenter can influence the pronunciation of participants (Hay, Nolan & Drager, 2006; Hay, Warren & Drager, 2006; Hay, Drager & Warren, 2009).

2.3.2 Hyperarticulation

Scholars who criticized earlier findings assume IN to be an experimental artifact caused by the formal set up in the laboratory. For example, Jassem and Richter criticised earlier findings of IN for their utterances “being far from spontaneous” (1989: 318) because of the metalinguistic nature of the carrier phrases (Hans sagte [] arg laut ‘Hans said [] very loudly’ (Charles-Luce, 1985)). According to Jassem and Richter, this type of presentation strongly favors “hypercorrectness or otherwise unnatural rendering” (1989: 318). Since several studies reporting IN effects did not use such metalinguistic carrier phrases (Port & Crawford, 1989; Röttger, Winter & Grawunder, 2011), this argument is not valid across the board.

Criticism is further related to the idea that participants in some of these experiments might have made an artificial effort to distinguish homophones (as also pointed out by Manaster Ramer, 1996: 487). This is especially likely when participants are asked to produce minimal pairs with many repetitions. Speakers are aware of the minimal pairs and could tend to hyperarticulate them. The results by Charles-Luce (1993) on the influence of semantic context on IN in Catalan seems to be in accord with this view (Manaster Ramer, 1996: 487). Most of the obtained durational differences ceased to be significant if the semantic context disambiguates the homophone.

However, a homophony distinguishing perspective would make the following prediction: the more homophones are presented to a participant in a short time, the larger the magnitude of the effect should be. This means that experiments with many fillers and distractor elements should be less likely to find IN or likely to find smaller IN effects. In contrast to this view, some of the studies that did find Incomplete Neutralization had many fillers (Dinnsen & Charles-Luce, 1984; Röttger, Winter & Grawunder, 2011).
2.3.3 Orthography bias

One concern that is always raised when IN is discussed is the role of orthography. Given that most of the above-mentioned studies used a reading task in languages that contrast the voicing status of obstruents through orthography, some authors concluded that IN was an experimental artefact of orthography (e.g. Fourakis & Iverson, 1984; Jassem & Richter, 1989; Warner et al., 2006). It is important to keep apart two different hypotheses regarding the relationship of orthography and IN: (1) orthography might be the direct cause of IN in an experimental setting, i.e. participants produce an artificial orthography induced contrast on the fly resulting in IN (‘hypercorrection’) (Fourakis & Iverson, 1984; Jassem & Richter, 1989); (2) orthography is one factor keeping underlying voiced and voiceless obstruents in neutralized positions apart from each other causing an inevitable permanent representational difference. In other words, speech production draws on both orthographic and phonological information (Manaster Ramer, 1996). Both claims will be discussed in the remainder of this section.

Fourakis and Iverson (1984) showed that when participants are asked to form a final devoiced verb form out of an auditorily presented non-neutralized form (meiden ‘to avoid’ → nied ‘avoid.PST.1+3SG’), no IN effect was obtained. In a reading task, however, the same participants showed a significant IN effect. Jassem and Richter (1989) prompted subjects auditorily with questions to which the target words formed one-word answers and found no significant IN effect in Polish when minimizing the role of orthography. However, in both studies numerical IN effects were obtained and might have been overlooked due to too small samples. A further argument for the impact of orthography on IN is the fact that minimal pairs of some studies reporting IN in part did not exhibit a phonemic difference between voiced and voiceless items, but only a difference in spelling as has been noted before by Fourakis and Iverson (1984) and Manaster Ramer (1996). This is the case for such words as seid ‘to be.PRS.2PL’ and seit ‘since’, for which Port and O’Dell (1985) and Port and Crawford (1989) reported IN effects even though many phonologists believe that there is no underlying difference with respect to the voicing of the final stop. Unlike Rad ‘wheel’, which provides evidence for the presence of an underlying voiced segment when it surfaces in intervocalic position in the morphologically related forms Rades or Räder (genitive and plural, respectively), seid does not participate in any relevant alternations. Recently,
Warner et al. (2006) showed that earlier findings on IN in Dutch (Warner et al., 2004) are primarily due to the orthographic representation of the contrast by showing that phonologically identical words were pronounced differently in accordance with voicing distinctions indicated by the orthography.

Port and Crawford (1989) recognized the potential influence of orthography and set out to address this problem by having a task in which speakers had to repeat what was being said to them by a native speaker of German. They cite the presence of IN effects in this task as showing the independence of the phenomenon from orthography; however, there are two major concerns. First, the German assistant who provided the utterance to be repeated could easily have been influenced by orthography because the stimuli were read to the participants. Because people accommodate to recorded nonwords (Goldinger, 1998), recorded individual syllables (Nielsen, 2005) and to computerized agents even if they do not feel sympathetic to these agents (Staum Casasanto, Jasmin & Cassanto, 2010), we know that convergence to an interlocutor’s speech patterns is a rapid and automatic process. Thus, accommodation of the participants to the assistant’s speech is a likely alternative explanation for the findings presented in Port and Crawford (1989). Second, whenever real words are used, the orthography bias cannot be avoided. This is because we know from a wealth of experimental studies that orthographic representations are automatically activated even when stimuli are presented in a completely auditory fashion (Perre, Midgley & Ziegler, 2009; Seidenberg & Tanenhaus, 1979; Ziegler & Ferrand, 1998). So there is ample evidence that orthographic representations are activated irrespectively of the task. This leaves us to conclude that if IN effects are induced by the orthographic contrast they should be omnipresent and cannot be avoided in a language that represents the lexical contrast orthographically.

What about IN in languages that do not represent the contrast in the spelling system? The evidence regarding the correlation of the presence of IN and an orthographic contrast is rather mixed. Dinnsen and Charles-Luce (1984) reported IN for Catalan (a language in which the neutralized contrast in question is not indicated by orthography), but the direction of the differences was different for the five speakers that were studied. Kopkalli (1993) and Wilson (2003) found no significant durational differences for word-final stops in Turkish that does not contrast the voicing distinction.
in its orthography. However, there are certain remarks to be made on the interpretation of the Turkish results, too. Wilson tested only three speakers on just a few test words. Inferential statistics over subjects were thus not possible (and not carried out). He did not provide the reader with sufficient information about the descriptive outcome of his acoustic measurement, therefore a critical evaluation of his results is not possible. Kopkalli tested five subjects and reported only on inferential results over items (30 minimal pairs). She measured the parameters duration of the vowel preceding the critical stop, voicing-into-the-closure duration, closure duration and aspiration. For the first three acoustic parameters, she found numerical IN effects: the mean vowel duration preceding underlying voiced stop was 3 ms longer than that of vowels preceding underlying voiceless stops. Voicing into the closure was 1 ms longer and closure duration was 3 ms shorter for underlying voiced stops. This IN effects were numerically present for all five speakers. For the parameter aspiration she found the tendency to produce more aspiration in underlying voiced stops consistent over all five speakers. In general, all of these numerical trends were not significant. However, for two speakers there was a significant interaction of underlying voicing and acoustic cue, indicating that some cues in these subjects differ significantly between underlying voicing categories. In sum, there seems to be numerical evidence for an IN effect in Turkish too, which might have been overlooked due to the small effect size and low statistical power.

Most importantly, Kim and Jongman (1996) investigated coda neutralization in Korean, whereby syllable final coronal obstruents (e.g. /t, tʰ, s/) are all neutralized to [t]. Even though the Korean spelling system does distinguish these segments orthographically, the authors did not obtain any evidence for IN of the contrast syllable finally. Finally, Yu (2007), on the other hand, found IN of alternating tones in Cantonese, which are not kept apart by orthography.

Ultimately, this debate seems not to be about a methodological concern regarding the stimulus material rather than about the wrong population tested. An experiment that presents stimuli to literate participants always obtains the orthography bias as a potential methodological confound. Thus hypothesis (1), the prevalent claim in the literature, seems to be superficially testable with illiterate people only. However, testing illiterate subjects will not necessarily settle this debate. Illiterate people participate in
communication situations with literate people every day, so they receive the same auditory input and in turn might accommodate to the perceived speech tokens resulting in representations similar to literate speakers.

Given the arguments above, the latter claim is impossible to test in a population exposed to a orthographic system that exhibits the contrast in the written system. However, the picture regarding the relation of IN and an orthographic contrast in the spelling system is rather pointing to an independence of IN and the orthographic contrast. We have to conclude that the hypothesis that German IN of the final voicing contrast is due to orthography does not seem to be a testable hypothesis. Researchers are left with trying to diminish its influence.

2.4 The settling of the debate?

Röttger and colleagues (2011) set out to demonstrate that IN is a robust phenomenon that cannot be explained away by methodological criticisms. To that end, they designed a production experiment on German final devoicing that circumvented the above problems associated with previous studies on IN. In their production experiment, which was inspired by the experimental set up of Iverson and Fourakis (1984), participants were asked to produce the “devoiced” singular form of a previously heard corresponding plural form (cf. (1) and (2)):

(1) Aus Köln kamen die Drude.
From Cologne come.PL.PAST DEF.3SG NONCE.PL
‘From Cologne came the druds.’

(2) Ein Drud wollte nicht mehr.
One.M NONCE.SG refuse.3SG.M.PAST NEG continue
‘One drud refused to continue.’

Crucially, they used pseudowords and introduced them as possible German nouns in the plural. Word frequency is obviously not a confounding aspect for pseudoword
production\(^1\). Additionally, the experimental design that they employed strongly drew participants’ attention away from the minimal pairs by having many fillers, and by having fillers that are difficult to respond to and thus require a lot of attention. In order to further disguise the real purpose of the task, they did not include any repetitions, and participants were told that they would perform a morphological task. The orthography bias was diminished by the use of a completely auditory design where at no point after having received the instructions, the participants had to read or write one of the test words. Note that participants could still have been influenced by orthography by automatically activating orthographic representations, as discussed above (Perre, Midgley & Ziegler, 2009; Seidenberg & Tanenhaus, 1979; Ziegler & Ferrand, 1998).

![Figure 2: Vowel durations for underlying voiced (white) and voiceless (gray) stops in the neutralizing context for all 16 speakers; top row: speakers 1 to 8, bottom row, speakers 9 to 16. Figure reprinted from Röttger et al. (2011: 1724).](image)

Finally, the statistical problems were addressed in various ways. First, they included more participants – crucially more than the proponents of IN in order to show a

\(^1\)Note, however, that word frequency of existing forms could come into play due to formal analogy to the pseudowords. As the mechanisms of analogy between pseudowords and existing words are not completely understood yet, one cannot rule out analogical interference with complete certainty.
sufficient attempt of proving the null (Frick, 1995). Then, they performed a mixed model analysis with both subjects and items as random effects, thus addressing the language-as-fixed-effect fallacy (Clark, 1973) and pseudoreplication (Winter, 2011). Finally, they corrected their results with respect to the amount of dependent measures that they analyzed. In general, all of these design and analysis aspects made it more difficult for IN effects to emerge in production, thus making any effect that they did obtain more robust.

They obtained a significant effect of duration of the preceding vowel: vowels preceding underlying voiced stops were produced on average 8 ms longer than vowels preceding underlying voiceless stops (cf. Fig. 2). The authors concluded that IN has to be taken as a relevant, genuine feature of the linguistic system that warrants explanation and is unlikely to be only induced by methodological errors.

In the next chapter, I am going to offer a promising account to explain the phenomenon.
Since IN does not seem to be a methodological artifact, the question arises as to what causes IN to surface? Some researchers have proposed that general cognitive properties of our mental lexicon could be the cause of IN effects. This idea is highly compatible with earlier experimental findings and makes promising empirical predictions. In the following, I will discuss an account that depicts IN as a cognitive artifact. Since it is grounded in recent ideas of exemplar-based models, I will take some time to describe these models in detail (for a recent overview and comparison between different exemplar-based accounts, see Ernestus & Baayan (2011)).

3.1 Exemplar-based models

It has been shown that lexical representations are rich in information, including detailed phonetic information of individual word forms (Brown & McNeill, 1966; Bybee 1994, 2000; Frisch, 1996; Goldinger, 1996, 1997; Palmeri, Goldinger & Pisoni, 1993; Pisoni, 1997). Recent studies have provided increasing evidence that these phonetic details are perceptible and used for word recognition (Davis, Marslen-Wilson & Gaskell, 2002; Hawkins & Nguyen, 2003). Thus, it seems to be an integral part of speech communication in both production and perception (see e.g. Hawkins, 2003). This evidence has led scholars to reject traditional modular feed-forward models, which assume minimal symbolic representations, and to turn to models that assume massive
Chapter III - The brain leaks through - Towards an explanation

storage of (redundant) concrete events. On a more abstract level, the lexicon is known to contain auditory and visual representations for a vast number of words. Importantly, it has been shown that people store complete and “ready-to-use” inflected forms in their mental lexicon (a.o. Alegre & Gordon, 1999; Baayen, Dijkstra & Schreuder, 1997; Butterworth, 1983; Bybee, 1995; Manelis & Tharp, 1977; Sereno & Jongman, 1997). Based on this evidence, there is ample justification to assume that the lexicon of a German speaker contains representations such as (1) as well as (2).

(1)  [ʁaːt]  Rad  ‘wheel’
(2)  [ʁæːðɐ]  Räder  ‘wheels’

These models, as opposed to traditional phonological accounts, do not assume a distinction between underlying and surface form. Rad is not stored as /ʁaːd/ with an underlying voiced alveolar stop, which is postlexically “devoiced” by a phonological rule. It is stored as a fully specified phonetic form [ʁaːt]. The exact nature of this representation is, however, a point of debate. While some researchers assume that lexical items are stored as massive exemplar clouds consisting of every perceived token of a particular word, others assume a more abstract lexical representation comparable to the phonemic transcription in (1) and (2). However, most exemplar-based models agree on certain assumptions about the organization of exemplars: exemplars are organized into a spreading activation network (e.g. Bybee, 2001; Hudson, 2007). Exemplars that share certain features are stored next to each other, in other words, the topological distance in the network is a function of formal and/or semantic similarity. Activation from one exemplar spreads blindly to its ‘neighbors’ and the strength of co-activation is negatively correlated with topological distance.

More detailed exemplar-based models assume the storage of concrete instances of memory traces. An exemplar-based model that incorporates more detailed exemplars in speech perception and production functions as follows (cf. Pierrehumbert, 2001): the language system is assumed to recognize input and generate output by an evaluation process across a lexicon of myriad distinct exemplars of each speech token activated in the long-term memory. Comparable to other exemplar-based models, this evaluation process is analogical, that is exemplars are inter-connected based on similarity. The
higher the frequency of a word, the higher the number of activated traces because of increased experience. A category is defined by a cloud of exemplars. Each new experience changes the entire exemplar system slightly through activation and range alterations (Goldinger, 1996, 1997, 1998; Hintzman, 1986; Nosofsky, 1986). Pierrehumbert (2001) states, that the production of a category label (e.g. [kaɾt]) results from a random selection of a category exemplar correlated with the probability of the base-activation level of an exemplar. So, even if there is an exact match to a stored word in memory “all similar activated traces create a ‘generic echo’, regressing toward the mean of the activated set.” (Goldinger 1997: 46). Thus the production target is the result of averaging exemplars within a region of an exemplar cloud (Pierrehumbert, 2002).

The study of phonetic convergence, the process in which a speaker acquires acoustic characteristics of the individual he is interacting with, has contributed a lot to the evaluation of different exemplar-based models. Early radical versions of exemplar-based models of speech perception like the one proposed by Goldinger (1996, 1997, 1998) predict that phonetic convergence is the automatic result of cognitive reflexes. By storing just heard instances of a particular form, the speaker shifts its category mean slightly towards the stored exemplar. The implicit imitation paradigm, in which subjects’ speech is compared before and after they are exposed to target speech, has shown that speakers shift their productions in the direction of what they have just heard. For example, Goldinger (1997) showed that subjects shifted their own F0 when they are asked to immediately repeat speech with manipulated F0. Shockley, Sabadini & Fowler (2005) extended these findings by showing a significant Voice-Onset-Time (VOT) imitation effect (in shadowing) for voiceless stops with artificially extended VOTs. These studies showed listeners’ sensitivity to variations in both suprasegmental and segmental phonetic dimensions.

Goldinger’s radical episodic trace model predicts that new exemplars of a lexical item have to compete with the effect of many thousands of existing episodes stored through regular exposure to language. This idea has been intensively discussed and criticized for its lack of abstraction (e.g. Cutler et al., 2010). It seems to be clear that the limited number of new exemplars added during a exposure phase is unlikely to be enough to change the whole representation of a word. Furthermore, several studies have
shown, that participants not only shift their pronunciation to the words they have been exposed to, but seem to generalize the just heard patterns to new words and other categories with the same sublexical units (e.g. McQueen, Cutler & Norris, 2006; Nielsen, 2011). Nielsen (2011) used an implicit imitation paradigm and exposed listeners to extended VOT target stimuli containing the stop /p/. Extended VOT stimuli led to speaker shift. Furthermore unheard words were introduced into the production list. The extended VOT patterns were generalized (i) to new words containing /p/ and, interestingly, (ii) to another voiceless stop category (/k/).

The assumption of a radical model that assumes storage of concrete exemplars only is not capable of explaining these findings. Both abstractions and detailed memory traces are required to explain the experimental facts. The above cited studies suggest that at least abstract prelexical representations of speech sounds such as “phonological features” or “phonemes” are likely candidates for abstractions made by the cognitive system. Consequently, mental representations of sound structures are likely to be multi-leveled. However, although several scholars have tried to argue in favor of these hybrid models (e.g. Abbot-Smith & Tomasello, 2006), they have not provided concrete formulations of how such models work.

### 3.2 Incomplete Neutralization as a cognitive by-product?

According to the PRODUCTION INTERNAL INTERACTION MODEL discussed by Baese-Berk & Goldrick (2009), “one production process influences the organisation and structure of processing within another production process” (ibd: 530f.). Through cascading activation multiple representations are activated that influence subsequent processes (e.g. Goldrick, 2006). For example, it has been shown that properties of lexical organisation have an impact on the phonetic realisation of an exemplar. Vowels of words in dense neighborhoods are more hyperarticulated than words in sparse neighborhoods (Munson, 2007; Munson & Solomon, 2004; Wright, 1997, 2004). Wright (1997, 2004) examined the vowel spaces of ‘hard’ and ‘easy’ words. Roughly defined, hard words have a dense lexical neighborhood with many high-frequency neighbors. Easy words on the other hand are words from sparse neighborhoods with primarily low-frequency neighbors. Wright demonstrated that vowels in hard vs. easy
words are realised with more extreme dispersion. These differences have been shown to be robust regardless of lexical frequency (Munson, 2007; Munson & Solomon, 2004). Now, in line with the above described view on the mental lexicon and the findings on lexical factors conditioning phonetic details, Incomplete Neutralization effects could be explained by co-activation of morphologically related exemplars (as is suggested by Ernestus & Baayen, 2006).

Baese-Berke & Goldrick (2009) demonstrated that hyperarticulation can be caused by minimal pair relationship. Activation allows multiple active lexical representations to influence processing not just of the target (e.g. Rad) but also of its formally and/or semantically related lexical neighbors that have been activated via co-activation from the target properties (e.g. Räder, Rades). These lexical neighbors activate not only phonological structures they share with the target (e.g. Räder and Rades provide additional support for the uvular fricative /ʁ/), but also phonological representations not shared with the target (e.g. the voiced alveolar stop /d/ or the open mid front vowel /æ/). Thus, these non-target representations compete with the target representation for selection. This competition is especially strong for segments with a high degree of featural similarity (e.g. /t/ and /d/). So let us assume speaker A wants to produce an instance of Rad. He will activate the target form [ʁÆt]. Additionally, close lexical neighbors are co-activated. The closest lexical neighbors (based on formal and semantic similarity) are morphological related exemplars. For the given example, it would thus be expected that the morphological related form [ʁÆ:t] (among others) becomes co-activated (Ernestus & Baayen, 2006). If some or most of the co-activated exemplars contain a non-neutralized segment that is voiced, these voiced forms could influence the motor commands used in speech production in subtle ways, leading to the small IN effects that has been observed.

Examples of paradigmatic analogical effects that ultimately lead to different productions are commonplace in historical linguistics (e.g. Trask, 2007). In the last two decades, this idea has been extended to synchronically observable phenomena (e.g. Benua, 1995; Burzio, 1994). One striking example for this is Yu (2007), who demonstrated that tonal near mergers in Cantonese are caused by or facilitated by

---

1 Following this argumentation, we would expect a systematic influence of the open mid front vowel /æ/ on the low vowel /a/, too.
interactions with their morphological neighbors, and that the phonetic realization of a derived tone may vary in the direction of its paradigmatic neighbor.

As there are no ‘paradigmatic neighbors’ for the pseudowords in Röttger et al.’s study (2011), how could this type of model explain the IN findings obtained? Within the analogy based account proposed by Ernestus & Baayen (2006), one could argue that participants established a representation of the just-heard plural pseudoword (e.g. [dʁuːdə]) that co-activated existing formally similar forms like [bʊːdə] ‘booth’ or [juːdə] ‘jew’. When participants produce the devoiced singular form (in this case [dʁuːt]), the co-activation of the just stored plural form and its neighbors containing the voiced stop [d] are activated as a by-product and interfere with the activation of the singular form containing the “devoiced” alveolar stop [t]. In the remainder of the paper, I will refer to this explanation as the CO-ACTIVATION HYPOTHESIS (henceforth CO-A, cf. (3)). Note, that CO-A predicts IN to occur every time a particular “devoiced” stop is produced.

(3) **CO-ACTIVATION HYPOTHESIS**
Incomplete Neutralization is a cognitive by-product caused by the simultaneous co-activation of competing lexical representations.

Now reconsider the hybrid models discussed above again. Based on stored exemplars, speakers are likely to form prelexical abstractions. Let us assume a prelexical category comparable to the phonological feature [VOICE] emerges out of the intervocalic voicing distinction of concrete exemplars. So, again, speaker A wants to produce an instance of [dʁuːt], he will co-activate the morphological related form [dʁuːdə] and its neighbors. It is not a big step then to hypothesize that an abstract prelexical category [VOICE] is co-activated, too. In the remainder of the paper, I will refer to this explanation as the PRELEXICAL CO-ACTIVATION HYPOTHESIS (henceforth PRELEXCO-A, cf. (4)). Note that CO-A and PRELEXCO-A are highly compatible.

(4) **PRELEXICAL CO-ACTIVATION HYPOTHESIS**
Through co-activation of a concrete exemplar, a prelexical categories associated to this exemplar are co-activated.
So IN could be explained as follows: through the experiment, participants are exposed to concrete phonetic manifestations of the *intervocalic voicing contrast* ($x_1, x_2, ..., x_i$; where $x$ stands for a concrete voiced-voiceless pair). These exemplars are stored and contribute to the overall exemplar dynamics of the prelexical voicing contrast intervocally ($X = /d/ \text{ vs. } /t/ \text{)}$. The phonetic properties of this category are associated to the phonetic properties of the neutralized contrast in final position ($Y (= \text{IN})$) (by highly overlapping exemplar clouds/co-activation). However, $x_i$ will modify the production of a concrete exemplar in final position ($y_i$) to a certain degree, because $x_i$ is its nearest lexical neighbor $y_i$.

### 3.3 Towards an answer

Beside the explanation proposed above and despite the arguments put forward in Chapter 2, one could still argue in favor of IN as an artifact of hyperarticulation to distinguish homophones or due to the orthography bias (e.g. Petra Wagner, personal communication). In the remainder of the paper, I will refer to this line of explanation as the **HYPERARTICULATION EXPLANATION (HYPER)**.

So, on the one hand, IN is assumed to be an artifact of the experimental setting. On the other hand, it is assumed to be a cognitive artifact due to lexical co-activation of morphologically related forms (Co-A).

Fortunately, the experimental design of Röttger et al. (2011) provides a first testing ground for these explanations. In their experiment they used a declination task in which participants heard a nonce word in the plural and had to produce the corresponding singular form. The plural-singular relation is assumed to obtain a very tight lexical connection: a plural form is semantically highly similar to the corresponding singular form, since the only semantic difference lies in the reference to the actual number of entities referred to. Moreover singular forms are formally highly similar to their corresponding plural forms due to the shared word stem. Because of the high similarity on both semantic and formal grounds, singular forms are assumed to be stored topologically very close to their plural equivalents. Thus according to the Co-A hypothesis, the form of the paradigmatic neighbor (plural form) should have an impact on the realisation of the target form (singular). In turn, manipulating the form of the
plural should have an impact on the realisation of the corresponding singular form due to lexical co-activation.

Having auditory presented plural forms that are clearly voiced or voiceless because of other cues (e.g. VOT, voicing, closure duration, etc.), it is possible to manipulate the duration of preceding vowels when the consonant is in intervocalic position. Note, that the duration of the preceding vowel was the only significant IN cue in Röttger et al.’s study. I will therefore refer to the vowel duration difference in final position as the IN effect even though other cues are might be relevant as well. The above discussed accounts make the following predictions regarding the phonetic manifestation of “neutralized” stops derived out of just heard intervocalic voiced and voiceless stops, respectively.

**HYPER** serves as the null hypothesis and does not predict any impact of the intervocalic manipulation on the IN effects, since IN is just a reflex of deliberately hyperarticulating the abstract categories [+VOICE] and [-VOICE].

**CO-A** assumes that the co-activation of the corresponding paradigmatic neighbors interferes with the production of the target, thus the form of the paradigmatic neighbor should modulate the interference. **CO-A** predicts the IN effect to be positively correlated with the vowel length difference in the intervocalic position. In other words, plural pairs in which the vowel duration difference is relatively large, should elicit relatively large IN effects; plural pairs in which the vowel duration difference is smaller, should elicit smaller IN effects. **CO-A** does predict a strict one-to-one mapping between the intervocalic and the final contrast. However, in line with PRELEX**CO-A** (cf. (4)), it is reasonable to assume a more abstract prelexical category of the intervocalic voicing contrast interfering with the concrete intervocalic voicing cues in the co-activated exemplars. In other words the local effect of the concrete co-activated exemplar is competing with the through life time experience established global prelexical pattern of the intervocalic contrast.

Furthermore, **CO-A** predicts a frequency effect: because the pseudowords are necessarily unknown to the participants, we have to assume a frequency of zero. Exposure to the same pseudowords should lead to increasing frequency that itself should lead to a stronger representation and a stronger connection to the paradigmatic form. **HYPER** does not make predictions regarding a frequency effect. In general,
assuming the null hypothesis, one would expect a decreasing trend due to the general observed reduction process in repeated items (Shields & Balota, 1991). The IN effect should be less hyperarticulated over repetitions and thus decrease. Co-A makes concrete predictions regarding frequency: the higher the frequency of the word, the higher its activation, resulting in more co-activation and in turn to a larger correlation between the intervocalic contrast and the IN effects. In the next chapter I report on an experiment that tests these competing predictions.
Chapter IV

Empirical validation

4.1 Method

4.1.1 Participants and experimental procedure

16 speakers participated in the experiment. All participants were native speakers of German without known hearing deficits (mean age: 27 years (SD = 5.81); 9 male, 7 female). All participants live in Cologne or in the area surrounding Cologne. Most of them grew up in this area. All aspects of the recording session were conducted in German by me.

Participants were seated in a well-illuminated sound-treated booth in front of a computer screen. They were given written instructions that stated that the experiment investigates phonetic regularities of German plural formation. In each trial, participants first heard a stimulus sentence such as (1) via headphones and then had to produce a corresponding sentence such as (2).

(1) Peter weiß nun, wie die Bauge sehen.  
Peter know.3SG now what DEF.PL NONCE.PL look like.3PL.  
‘Peter knows now what the NONCE look like.’

(2) Denn nur der Baug sieht so aus.  
As only DEF.3SG.M NONCE.SG look.3SG like V-PART  
‘As only the NONCE looks like this.’
The critical phoneme of this example is /g/, which appears in a voiced and non-neutralizing context in stimulus (1) but which had to be pronounced in a neutralizing context in the response (2). The instructions introduced the pseudowords as artificial words and asked the participants to treat them like German words. The plural stimuli were produced with following intonational pattern: in terms of GToBI (Grice, Baumann & Benzmüller, 2005), an L+H* was located on weiß ‘know.3sg.’ and a high small boundary was obtained after nun ‘now’ (H-). The nuclear accent of the subordinate clause realized with an !H* or L* was located on the first syllable of aussehen ‘look like’. The prenuclear accent realized with an L+H* was located on the first syllable of the target word (cf. fig 3).

Figure 3: Acoustic waveform, spectrogram, $f_0$ contour, and GToBI annotation of the auditory stimulus Peter weiß nun, wie die Bauge aussehen ‘Peter knows now what the Baugs look like’.

The experiment was controlled using Superlab 2.04 (Abboud, 1991). At the beginning of each trial, a cross appeared in the centre of the screen (+) and participants heard the plural sentence through headphones. After an inter-stimulus-interval of 500ms three question marks (???) appeared on the screen. Participants were now asked to produce
the corresponding singular sentence. The experiment was self-paced and there were no time constraints. An empty screen appeared for 500ms after each trial.

Before the actual experiment, participants listened to 16 demonstration stimuli to introduce them to the operational sequence of the experiment. They heard the plural sentence and a modeled singular response sentence. The singular forms were produced with a nuclear accent on the target word manifested through an L+H* (cf. Fig. 4).

None of the demonstration items were potential critical items (see Stimuli section, cf. Appendix B) in order to avoid a possible IN bias. If the demonstration stimuli, which were read out, had contained devoiced obstruents, they could have already contained IN cues, thus leading to an IN model for the participants. The demonstration items were necessary for two reasons: on the one hand, the demonstration session was introduced to avoid disyllabic and non-neutralizing responses. On the other hand, it was used to

\[ \text{E.g. } \text{Denn nur der Bauge sieht so aus; some German words are disyllabic in both the plural and the singular.} \]
prime participants to produce the target word with a particular intonational pattern, i.e. realized in a prominent prosodic position, that is in nuclear position.

After the demonstration session, participants were given the opportunity to perform eight practice trials (cf. Appendix C).

The actual experiment was divided into four blocks. After each block, there was an obligatory break of at least ten seconds. On average, the entire experiment (including the instruction and the debriefing) took about 30 minutes.

4.1.2 Speech material

The experimental items consisted of 48 pseudoword pairs such as (3)-(4) (cf. Appendix A):

(3) \([\text{viːb} \dot{\text{a}}] \) vs. \([\text{viːp} \dot{\text{b}} \text{a}]\)
(4) \([\text{gaːg} \dot{\text{a}}] \) vs. \([\text{gaːk} \dot{\text{b}} \text{a}]\)

All stimuli followed German phonotactics and were stressed on the first syllable. There were 24 stimuli pairs with labial and 24 with velar stops, each of which followed one of the vowels /i, e, a, aɪ, o, u/. Stimuli were balanced for vowel quality. Each experimental item was introduced as a masculine noun inflected for plural. Plural inflection was indicated through the regular plural marker for masculine nouns (/-\(\text{a}\)/), the plural determiner /\(\text{di}\)/, and number agreement on the verb. There were no fillers, thus the contrast between the corresponding members of a minimal pair was very obvious for the participants.\(^2\) This might lead to an additional enhancement of a potential IN contrast (Jassem & Richter, 1989).

The 48 stimuli pairs were spoken by a trained native speaker of German (male) along with the demonstration and practice items in a sound-treated booth recorded with an AKG C420 III microphone (44.1 kHz (16bit)).

\(^2\) However, in the postexperimental interview, only two participants explicitly mentioned that they were aware of minimal pairs distinguished through voicing. Since the interview did not explicitly asked for this information, no valid evaluation of the minimal pair awareness can be made.
4.1.3 Acoustic analyses of auditory stimuli

In order to control that the voicing distinction was indeed present in the input stimuli, I performed an acoustic analysis with Praat 5.2 (Boersma & Weenink, 2009). I analyzed the duration of the vowel preceding the critical stop, the closure duration, and the aspiration duration. For each of these measures, I performed an item-based paired t-test. I corrected for performing three tests on three dependent measures by means of Dunn-Šidák correction.

The vowel preceding the critical stop was on average 24ms shorter for voiceless stops ($t_2(47)=9.07$, $p<0.0001$). The closure duration was on average 21ms longer for voiceless stops ($t_2(47)=7.87$, $p<0.0001$), as well as the aspiration duration, which was on average 47ms long in case of voiceless stops ($t_2(47)$ against zero $=26.99$, $p<0.0001$). All of the voiced stimuli had voicing during the closure. To summarize, there were large differences for the duration of the preceding vowel, the closure duration and the voice onset time, as well as differences in voicing during the closure. This means that beside vowel duration there were at least three robust cues for participants to distinguish between voiced and voiceless stimuli.

4.1.4 Stimuli manipulation

I calculated the percentage of the vowel duration preceding the voiceless stop in comparison to the vowels preceding the voiced stop for each minimal pair. The mean difference was ~16%. This value was taken as a baseline. Based on calculated percentages, I assigned minimal pairs to four conditions. Conditions were balanced for place of articulation of stops (labial vs. velar) and vowel quality. I tested whether the conditions differ from each other regarding the acoustic parameters closure duration and aspiration. To that end, I performed an one-way ANOVA for each parameter. The analysis confirms that there was no difference between the conditions ($F(3,11) \leq 1.228$, $p \geq .31$).

I manipulated the vowel durations via using PSOLA resynthesis in Praat 5.2 (Boersma & Weenink, 2009). This change preserves pitch contour and vowel quality characteristics while resulting in natural-sounding syllables. Each minimal pair of the first condition obtained a difference in vowel duration of 32% (henceforth ENHANCED), that is vowels preceding underlying voiced stops are 32% longer than vowels preceding...
voiceless stops. Each minimal pair of the second group was manipulated so that the voicing distinction realized through vowel duration was 16% (henceforth **NORMAL**) reflecting the regular voicing distinction pattern observed in our recordings (½ * **ENHANCED**). Each minimal pair of the third condition was manipulated so that the voicing distinction realized through vowel duration was neutralized resulting in 0% (henceforth **NEUTRALIZED**). Each minimal pair of the fourth condition was manipulated so that the voicing distinction realized through vowel duration was reversed resulting in a mirror image of **NORMAL** (henceforth **REVERSED**). In other words after the manipulation the fourth group contained stimuli with 16% longer vowels preceding underlying voiceless stops than preceding underlying voiced stops (cf. Figure 5).

![Stimuli](image)

**Figure 5:** Manipulation conditions.

### 4.1.5 Stimuli norming

I examined the perceptual robustness of the voicing distinction in the manipulated forms by conducting a norming study. Three native speakers of German (one male, two female; mean age: 24) were asked to decide whether the presented stimuli were voiced or voiceless in a forced-choice experiment. The norming study confirmed that the
voicing contrast of the critical stimuli is very easy to perceive: participants did not make any errors in identifying the voicing of a stop. Even though I manipulated one perceptual cue for the voicing distinction, participants seem to rely primarily on other cues like VOT, closure duration and voicing of the closure.

4.1.6 Stimuli presentation

All stimuli were randomized for each participant. The actual experiment was divided into four blocks. The first two blocks contained all 48 critical pairs balanced for place of articulation of the stop, vowel quality and condition. A subset of these items was repeated twice in block three and four containing only the items of the enhanced and reversed condition (the most extreme stimuli with respect to vowel duration differences). In other words, stimuli of the enhanced and reversed condition occurred three times in the course of the experiment. Corresponding members of a minimal pair in the first two blocks were separated by one block (so by at least 48 items). Corresponding members of a minimal pair in the last two blocks were separated by at least 24 items.

4.1.7 Acoustic analyses of production data

The speech was digitized at a sampling rate of 44.1 kHz (16bit), and all measurements were made using Praat 5.2 (Boersma & Weenink, 2009). The experiment includes only the vowel duration of preceding vowels as dependent measures. I labeled all stimuli blindly. If the end of the preceding segment was voiceless (e.g. [ʃtaʊ̯k], onset of the vowel was defined as onset of voicing. If the vowel followed a voiced stop (e.g. [ɡoːp]), onset of the vowel was defined as the end of the burst. A sudden discontinuity in the spectrogram was taken as the boundary for vowels following nasals (/m/ and /n/), palatal approximants (/j/) and liquids (/l/ and /ɾ/) (e.g. [muːp], [jiːt], or [fɪrət]). The end of the vowel and the beginning of the closure were defined as the end of the second formant of the vowel, which usually coincided with a sudden drop in amplitude of voicing.
Chapter IV - Empirical validation

4.1.8 Statistics

All data were analyzed using R and the packages lme4 (Bates & Maechler, 2009) and languageR (Baayen, 2009). Data was analyzed by using linear mixed effects models (Baayen et al., 2008). In order to avoid the language-as-a-fixed-effect fallacy (Clark, 1973) and pseudoreplication (Winter, 2011), I used both Subjects and Stimuli pairs as random effects (Baayen, Davidson & Bates, 2008). To check if the IN effects are dependent on the manipulation of conditions, I performed an overall model (Omnibus Model) on all data points, which take the vowel duration as a dependent variable. I included the ordinal factor Occurrence and the two-way interaction of the categorical factor Underlying Voicing with the levels “underlyingly voiced” and “underlyingly voiceless” and the ordinal factor Condition with the levels “Enhanced”, “Normal”, “Neutralized”, and “Reversed” as fixed effects. Because the factor Occurrence was only multi-leveled for a subset of the data, this model was not able to test for the interaction between Occurrence and Condition. To test for the relevant interactions, I performed a subset model (Subset Model) on the data points of the ENHANCED and REVERSED condition. I included the three-way interaction of the categorical factor Underlying Voicing, the ordinal factor Condition with the levels “1”, and “4”, and the ordinal factor Occurrence with the levels “1”, “2”, and “3” as fixed effects. Both models included the categorical factor Place of Articulation with the levels “bilabial” and “velar” as well as the categorical factor Vowel Quality with the levels “i” “e” “a” “au” “o” and “u” as additional predictors. These were significant predictors of vowel duration, but this was expected: these factors were included mainly as controls, to insure that undue explained variance was not being attributed to the predictors of interest. Because mixed models do not have degrees of freedom I will only report p-values.

I checked for normality and homogeneity by visual inspection of plots of residuals against fitted values. Moreover, I performed model validation: results will not be reported if a null model (without fixed effects) does not differ significantly from the test model (with fixed effects) in a likelihood ratio test. Throughout the paper, I present p-values based on Harald Baayen’s Markov Chain Monte Carlo (MCMC) simulation function.
4.2 Results and discussion

Because I used pseudowords that were necessarily unknown to my participants, I had to exclude many responses that were either incorrectly remembered or produced with a lot of hesitation. When an item was excluded, the corresponding member of the minimal pair was consequently excluded, too. In sum, I had to exclude 7.2% of the productions resulting in 2852 items considerable for the analyses.

In general, there was a small vowel duration difference between voiceless and voiced stops that went in the predicted direction: vowels preceding underlying voiced stops were slightly longer than those preceding underlying voiceless stops (4.1 ms, cf. Fig. 6).

Figure 6: Vowel durations for underlying voiceless (white; mean=123.72 ms, SD=31.16 ms) and voiced (gray; mean=127.86 ms, SD=31.68 ms) stops in the neutralizing context.

Figure 8 depicts the duration differences of the vowel preceding final underlying voiced and voiceless stops for each individual speaker, showing a pronounced consistency

---

3 Cf. Appendix D. Tables show the output of the models, fitted in R (R Development Core Team, 2008) using functions contained in the languageR package. The model estimates (in column 1) are extremely similar to the mean estimates across 100,000 Markov Chain Monte Carlo (MCMC) samples (column 2).
across subjects. Only two speakers were found not to have shorter vowels before underlying voiceless stops (speaker 4 and 7 in Fig. 7). The present results confirm earlier findings of IN. Vowels preceding underlying voiced stops are longer than vowels preceding underlying voiceless stops in final position.

![Figure 7: Vowel durations for underlying voiceless (white) and voiced (gray) stops in the neutralizing context for all 16 speakers; left: speakers 1 to 8; right: speakers 9 to 16.](image)

**4.2.1 Omnibus Model (first production of all stimuli)**

There was a significant main effect of *Underlying Voicing*, indicating an IN effect (p<.0001). An explanation for the small numerical magnitude could be rooted in the hypothesized effects of the vowel manipulation. The two-way interaction of *Underlying Voicing* and *Condition* was indeed significant (p=.0391) indicating that the IN effect is modulated by the manipulation of the intervocalic contrast.
Let us consider the proportional IN effects for each of the four conditions (cf. Fig. 8, positive values represent IN effects in the predicted direction, i.e. vowels preceding underlying voiced stops are longer than vowels preceding underlying voiceless stops). Descriptively, positive proportional values are obtained for all four conditions, pointing to IN effects in all four conditions. As the family-wise error rate is already controlled for by the significant result of the *Omnibus Model*, I tested if the conditions elicited robust IN effects in their own terms. I constructed a model with *Underlying Voicing* as the fixed effect for each condition separately\(^4\). The effect of *Underlying Voicing* was indeed significant for all conditions (p<.0001, p=.0002, p=.0204, and p=.0130, respectively). Thus, the significant interaction of *Underlying Voicing* and *Condition* has to be taken as evidence for different degrees of Incomplete Neutralization.

---

\(^4\) For reasons mentioned in §4.1.8, *Vowel Quality*, *Place of Articulation* and *Occurrence* were also included as fixed effects.
Furthermore, there seems to be a decreasing tendency of this effect from the enhanced condition to the reversed condition. This stepwise trend shows the positive correlation between the intervocalic cues and the IN effects: stimuli with an enhanced intervocalic effect obtained larger IN effects (mean = 4.51%, SD = 15.65%) than stimuli with normal intervocalic effects (mean = 3.64%, SD = 15.14%). Stimuli with neutralized intervocalic vowel duration differences obtained even smaller IN effects (mean = 3.53%, SD = 17.28%), followed by even smaller IN effects for the reversed condition (mean = 1.94%, SD = 15.34%).

There was a significant main effect of Occurrence (p=.002). Interestingly, repetitions led to increasing vowel durations (1st occurrence (125ms), 2nd (126ms), 3rd (128ms)). That is surprising because we would expect a general decreasing trend due to the general observed reduction process of repeated items (Shields & Balota, 1991).

4.2 Subset Model (repeated stimuli)

In the Subset Model, there was a significant main effect of Underlying Voicing, too, indicating an overall IN effect (p<.0001).\(^5\) Crucially, the two-way interaction of Underlying Voicing and Condition was significant (p=.004) (cf. Figure 9) providing further evidence for the obtained effects in the Omnibus Model: the enhanced condition elicited significantly greater IN effects (4.5% (6ms)) than the reversed condition (1.9% (2.3ms)).

These findings are perfectly in line with Co-A’s predictions (cf. § 3.3). After exposure, concrete phonetic manifestations of the intervocalic voicing contrast were stored \((x_1, x_2,...,x_i)\). These exemplars contributed to the overall exemplar dynamics of a prelexical voicing category \((X)\). The phonetic properties of \(X\) are associated to the the phonetic properties of the neutralized contrast in final position \((Y (=IN))\). Since there was a global IN effect in all conditions, it is reasonable to assume that the properties of \(X\) dictated the response to a concrete form \(x_i (y_i)\). This explains the global IN effect in the experiment. However, \(x_i\) did modify \(y_i\) to a certain degree. This can be interpreted as a local effect and explains the interaction of Condition and Underlying Voicing. HYPER fails to explain these patterns.

\(^5\) Note that this is not surprising since this analysis is dependent on the first model due to shared data points.
The obtained significant main effect of *Occurrence* of the *Omnibus Model* was further confirmed (p<.0001)\(^6\). However, the interaction of *Condition* and *Occurrence* was significant as well (p=.0417). The increasing trend of repetitions seems to be driven by condition 1 (enhanced). Vowel durations of condition 1 increase slightly when repeated (123ms, 124ms, 126ms), while condition 4 (reversed) shows no increase (128ms, 127ms, 129ms).

Neither the two-way interaction of *Underlying Voicing* and *Occurrence* (p=.1113) nor the three-way interaction of *Underlying Voicing, Condition*, and *Occurrence* was significant (p=.117). So inferentially, the IN effects and the modulation of the effects through condition seems to be unaffected by increasing frequency. However, due to the fact that all of the discussed effects are rather small, let us consider the obtained patterns descriptively (cf. Fig. 10 and 11).

\(^6\)Note, again, that this is not surprising since this analysis is highly dependent on the first model due to shared data points.
Due to the high variance, it is hard to detect any insightful patterns concerning the occurrences in the boxplot depiction. In Figure 10, the data are plotted in barplots. In the second occurrence speakers seem to enhance the IN effect in both conditions. Furthermore, the patterns look very similar to the first production: the enhanced condition elicited greater IN effects than the reversed condition (4.89/0.72% (6.6/1.1ms) and 6.85/1.8% (8/2.2ms), respectively). However, in the third occurrence the effect sizes seem to approximate each other (2.59/3.18% (3.4/3.4ms)). Note that the frequency based explanation of Co-A predicts a different pattern: the higher the frequency of the word, the higher its activation, resulting in more co-activation and in turn in a greater correlation between the intervocalic contrast and the IN effect. That is not the case here.

Figure 10: Mean proportional IN effects for enhanced (black bars) and reversed condition (gray bars) over occurrences. For comprehensibility, no variance is given here (SD over 14% for all conditions).
In fact, when one has a look at the effects in relation to each other, it becomes clear that the effects are approximating each other linearly. In the first production, the enhanced IN effect is almost seven times as large as the reversed IN effect. In the second occurrence, the difference is reduced to a 1-to-4 ratio. Finally, in the third occurrence the enhanced and reversed IN effects approximate each other (the ratio is close to 1.0) (cf. Fig. 12).

Even though there is no inferential evidence for an interaction of Underlying Voicing and Occurrence, I would like to offer an explanation for the descriptive patterns obtained. The patterns could be explained by the nature of the presentation modus. It has been shown that listeners can easily be trained to shift phonemic category boundaries after a limited amount of exposure to a training set (e.g. Norris, McQueen & Cutler, 2003). However, listeners seem to apply the category boundary modulation only to contrasts uttered by the same talker (Davis et al., 2005; Eisner & McQueen, 2005,
Kraljic & Samuel, 2006; Maye, Aslin & Tanenhaus, 2008). Cutler et al. explain this phenomenon as follows:

“Once the prelexical system had learned about a talker idiosyncrasy which affected, for example, a single phoneme, that learning would automatically generalize across the vocabulary […]” (2010: 93).

According to this view, speakers build up an abstract representation of the speaker idiosyncrasy, which is based on the perceived exemplars. Since participants in the present study were exposed to just one speaker voice, after sufficient input, participants might have built up an abstract representation of the speakers idiosyncratic pronunciation patterns that averages the variability. This abstract representation (X) is an averaged reflection of the perceived exemplars ($x_1, x_2, ..., x_i$). Note that over the stimuli set, the average vowel duration difference between underlying voiced and voiceless stops was 16% in favor of voiced stops (reflecting the regular pattern). In other words, the stimuli design was skewed. This generalized voicing contrast representation at the end of the experiment (X), in turn, affected the IN effects ($y_1, y_2, ..., y_i$) to approximate each other resulting in an invariant overall Y. I will refer to that as the SPEAKER NORMALIZATION EXPLANATION (SNE). One could go one step further and propose that while the SNE accounts for the representation of the intervocalic voicing contrast, a similar normalization could be at work for the IN effects in respect to the speakers’ own productions. Speakers start to produce the IN effects in a way that is highly dependent on the just heard (idiosyncratic) intervocalic contrast. In the course of the experiment, they normalize their own idiosyncratic productions. Obviously, these ideas are not mutually exclusive. However, it seems to be hard to disentangle them empirically. It is worth noting that this explanation is rather post-hoc. Nevertheless, it is perfectly compatible with the above described model of linguistic representation (cf. §3).

Fortunately, this account is not even incompatible with the frequency based explanation. While the frequency based explanation is bound to the repeated occurrence of the same items, SNE is mainly dependent on the exposure duration to a particular speaker. This in turn makes testable predictions: if SNE is right, IN effect differences between conditions decrease over exposure time in an experiment without repetitions.
So if repetitions were produced by a different speaker, speaker normalization would not occur and the predicted frequency effect should be observable. Thus I reasoned that the lack of a frequency effect in our data can be due to other effects, which the experimental design was not controlling for. SNE has even further implications: the rather weak impact of vowel duration manipulation obtained in the results might be partially due to this normalization process. SNE might have masked the genuine impact of the manipulation. Again replications using different speakers might shed light on these issues.

Finally, a not so trivial remark on this interpretation has to be made: although the given explanation is compatible with our present knowledge of speech processing and might be intuitively plausible to some of you, it has to be kept in mind that I am interpreting numerical tendencies here. These numerical patterns show no evidence that they are statistically different from random noise. Till further data are collected, these interpretations have to be taken carefully.

4.3 Summary

To sum up, in a declination task participants heard pseudowords in their plural form containing voiced and voiceless stops in intervocalic position. Their task was to produce the corresponding singular form containing the neutralized voicing contrast word finally. In the plural forms, one cue for intervocalic voicing was acoustically manipulated to create a continuum of different durations of the preceding vowel.

The results showed an IN effect, i.e. vowels preceding underlying voiced stops in a neutralizing context are longer than those preceding underlying voiceless stops. This size of the contrast was very subtle and was comparable to IN effects obtained in previous experiment for German (Port & O’Dell, 1985; Röttger, Winter & Grawunder, 2011) and other languages (Dinnsen & Charles-Luce, 1984; Dimitrieva, Jongman & Sereno, 2010; Slówicka & Dinnsen, 1985; Warner et al., 2004). The magnitude of the effect was further modulated by the manipulation of the stimuli. However, although there was a positive correlation between the intervocalic voicing difference and IN, there was no one-to-one mapping. Even the stimuli exhibiting neutralized or a reversed (thus unnatural) vowel duration difference produced – albeit small –IN effects in the
natural direction. These findings are in line with the co-activation hypothesis and the prelexical co-activation hypothesis discussed in §3.3.

According to the proposed account, the overall IN effect is caused by the co-activation of a global prelexical voicing category, the exemplar specific correlation is caused by the co-activation of concrete neighbors, namely the just heard plural forms. The expected interaction effect of IN and frequency was not obtained, i.e. the correlation of intervocalic contrast and IN effect did not increase through repetitions. A possible explanation for the lack of a frequency effect might lay in speaker normalization processes (e.g. Cutler et al., 2010). Since participants were exposed to just one speaker voice, after sufficient input, they might have built up an abstract representation of his idiosyncratic pronunciation patterns. This abstract representation might have masked exemplar-specific effects like frequency. For future research, one should take this potential effect into account. In particular the obtained results should be replicated with an intersubject design, i.e. one group of participants had to respond to stimuli of the reversed condition only, and another group had to respond to stimuli of the enhanced condition only. This design would be balanced and could give an answer to the question whether the overall IN effects obtained in the present experiment are due to the unbalanced design and the impact of SNE or a genuine prelexical co-activation effect. Further, one has to disentangle the frequency effect from SNE by comparing the results of an experiment with repetitions spoken by a single speaker to a design with repetitions spoken by different speakers.
Chapter V

Putting the picture together

The exemplar-based view ascribes the effects observed in IN experiments to general cognitive mechanisms such as memorizing, categorizing and analogy. It is highly compatible with our current knowledge of language processing in both production and perception. It further has the beauty of accounting for the inter-speaker and contextual differences that are found in previous experiments. Given that individual experiences form the basis for further abstractions, this account can also easily and parsimoniously explain the individual differences observed in other studies. This account thus also circumvents the problem of accounting for individual differences phonologically, e.g. Dinnsen & Charles-Luce (1984) who consider speaker-specific implementational rules. Furthermore, it allows for some amount of IN to occur in real-life. Depending on circumstance, some occurrences of words will exhibit small production differences. Through lexical co-activation and through re-occurrences of words with these small production differences, a small “semicontrast” could potentially persist in a population of speakers and a population of words (cf. §5.1). In light of the proposed account the rather weak perceptual salience of this contrast is not that surprising anymore, as will be discussed in §5.2 in more detail. As opposed to the attempts to incorporate IN in traditional phonological models, the suggested model makes clear empirical predictions both typologically as well as experimentally, which are highly dependent on languages specific organization of the lexicon. These will be discussed in §5.3.
5.1 To be or not to be? The representation question

We have identified two possible different mechanisms causing IN in production: the cognitive by-product of lexical analogy (CO-A) and the potential hyperarticulation either caused by homophony distinguishing or an orthographic induced contrast (HYPER). For the sake of discussion and beside both the above discussed arguments against the latter (cf. §2) and the present experimental results (cf. §4), let us assume HYPER is a valid explanation. The question arises as to whether speakers hyperarticulate on the fly leading to spontaneous IN effects that are potentially caused by unnatural lab speech conditions and thus are not represented in the speakers mind, or if IN is indeed based on a representational difference. The same question holds for CO-A. Since this account has to assume a fully automatic cognitive process that cannot be avoided even in natural speech, IN would have to be accepted as a genuine effect. However, if the co-activation of lexical neighbors leads to different representations of underlying voiced and voiceless obstruents in final position, is not clear yet. In line with the above described exemplar-based models, I would like to argue that both HYPER and CO-A cannot be assumed to be just on-the-fly processes without any influence on the cognitive representation.

It was stated that speakers might hyperarticulate the contrast between underlying voiced and voiceless obstruents in final position in the laboratory (HYPER). This claim implicitly assumes that speakers are able to hyperarticulate the distinction in the first place. Lets consider following scenario: in the moment a speaker A produces this hyperarticulation on the fly, a listener B will store this hyperarticulated exemplar, which, in line with exemplar-based assumptions, changes the entire category system slightly. Thus, the production of hyperarticulated final obstruents of different underlying voicing status in certain contexts could lead to two slightly different representations. This, of course, is also true for CO-A, except that the contrast should be even more strongly represented. Since the co-activation of lexical neighbors is predicted to happen automatically, categories should be permanently updated with slightly different exemplars.

In the case of final devoicing in German obstruents, I assume following scenario. The exemplar cloud of two categories (underlying voiced and voiceless stops) overlap
substantially but the category membership remains distinct from each other. This could be due to several factors: an ambiguous token in natural communication is sufficiently disambiguated by the morphosyntactic and semantic context. It is not ruled out, though, that the orthographic reflection of the contrast in the German spelling system could help the system keeping the categories apart too. The production of the two categories (e.g. /d/ and /t/ in final position) proceeds by creating a target based on the average of some number of exemplars. Because of highly overlapping clouds, exemplars within the region of the partially overlapping exemplar clouds are produced as well. The mean production difference between these two categories are very small, but will be robust. This exemplar based idea is consistent with Yu’s (2007, 2011) interpretation of near mergers. In a nutshell, he assumes that near mergers are a “by-product of extensive overlapping of the exemplar clouds of two or more categories” (Yu 2007: 190). In that model, IN would be nothing different than a near merger.

When two or more exemplar clouds of different categories overlap extensively, listeners are predicted to have difficulty in reliably identifying tokens falling within the combined exemplar space. The production difference between near-merged tokens is predicted to be small, due to the proximity of exemplars being drawn for calculating the production target. (Yu 2007: 209)

### 5.2 Incomplete Neutralization in perception

Since we have pointed to a possible explanation for the phonetic manifestation of IN in production, the results of perceptions studies on IN become less remarkable. Listeners have been shown to be sensitive to IN, since they can tell apart above chance level the phonetic realizations of words underlyingly ending in voiced and voiceless obstruents (Kleber, John & Harrington, 2010; Port & Crawford, 1989; Port & O’Dell, 1985; Röttger, Winter & Grawunder, 2011). In natural communication, the listener is able to discriminate the instances of putative minimal pairs (Rad vs. Rat) based on the given morphological, semantic and pragmatic context (note, for example, that Rad is accompanied by the article inflected for neutrum (das Rad), while Rat is accompanied by a masculine article (der Rat)). However, in a forced choice paradigm, confronted
with isolated words and in absence of any disambiguating contexts, the listener has to make a decision based on minute acoustic differences.

Ernestus & Baayen (2006) argue for the same process of analogical co-activation in perception of IN. Let us assume, that speaker A hears [ɾæːt] and both word forms Rad and Rat with their corresponding nearest neighbors were activated (i.e. [ɾæːdə] and [ɾæːtə]). They assume that “[t]he activation levels of these neighbors are co determined by the size of the mismatch between the acoustic signal and the auditory representations of the words.” (ibid: 47). If the input stimulus has been produced with IN cues for a underlyingly “devoiced” stop, these subtle cues could lead to probabilistically more activation of the neighbors with the voiced stop because “The final activation levels of the neighbors co-determine the interpretation of the acoustic signal.” (ibid).

Accepting the fact that IN in production could emerge through co-activation of lexical neighbors leads us consequently to the conclusion that speakers systematically produce Rat and Rad in slightly different ways (as discussed above). Model theoretically, we do not need the production explanation reduplicated for the perception of IN. The listener just has to hear this slightly different forms and store the instances as exemplars of the intended lexical item. These exemplars influence the exemplar cloud of the intended forms and lead to slightly different representations. When listeners perceive a word form, auditory representations that best match the just heard form are activated. Let us assume again, speaker A hears [ɾæːt]. Both word forms Rad and Rat are activated, because of highly similar acoustic forms. However, one of these forms is more similar to the just heard acoustic event than the other based merely on slightly different representations. This is highly consistent with Warner et al.’s findings (2004) that speakers who produce a bigger IN effect are more likely to perceive the contrast in the perception experiment. Nevertheless, it is reasonable to assume the analogical process to be bidirectional, that is influencing both perception and production. The results by Warner et al. (2004) on closure duration in Dutch speak to an analogy-based account. In their production experiment, they found no difference of closure duration in the neutralizing context, however, speakers used closure duration as a cue to a final voiced or voiceless stop in a perception experiment. Warner et al. (2004: 269f.) proposed an explanation that is very much in line with analogy: listeners locate the acoustic dimension that is manipulated and base their choice of /t/ vs. /d/ responses on a
comparison with the acoustic dimension to the dimension in the non-neutralizing context. This seems to be the only way of explaining that the direction of the effect (longer closure durations were more likely to be interpreted as /t/) is consistent across subjects and matches the direction of the difference in a non-neutralizing context.

The offered explanation is compatible with most facts of IN. Even though listeners were able to discriminate underlying voiced from voiceless forms better than chance, the judgment accuracies are very low (cf. Port & Crawford, 1989; Port & O’Dell, 1985; Slowiascek & Szymanska, 1989, Röttger, Winter, & Grawunder). Because of highly overlapping representations, this low accuracy is not surprising.

5.3 Incomplete Neutralization and lexical organization

The proposed account predicts IN to be tightly connected to and dependent on properties of the mental lexicon. This leads to clear typological and experimental predictions. In line with our knowledge of lexical organisation and processing, we expect words that have very frequent morphological neighbors with voiced stops in intervocalic position to exhibit stronger IN effects than words with very infrequent morphological neighbors with voiced stops in intervocalic position. Furthermore, the number of morphological neighbors with voiced stops should matter: we expect words that have many morphological neighbors with voiced stops to exhibit stronger IN effects than words with less morphological neighbors. Note that the lexical neighbors do not have to be necessarily morphological related. Semantically or phonologically close neighbors could trigger the IN effect, too. However, since morphological neighbors are both semantically and formally very similar to a target form, IN effects should be triggered particularly by those forms. These are empirically testable prediction that sets the proposed CO-ACTIVATION HYPOTHESIS apart from traditional phonological explanations discussed in §2.2.

Since the proposed account is based on general processing properties of lexical organization, it predicts all morphologically sensitive neutralization processes to be incomplete. This is in fact a very strong claim, which could be easily falsified by cross
linguistic data. I would like to argue that one has to consider IN cross linguistically in light of the lexical organisation of the language under investigation.

The case of Turkish discussed in §2.3.3 can barely be considered as a falsification of Co-A, since genuine IN effects might have been overlooked due to small obtained effect size and low statistical power. To my knowledge, there is no serious challenge to the prediction that IN occurs in all experimentally investigated cases of final devoicing yet. However, a genuine challenge to my proposal is the complete neutralization of Korean manner of articulation features reported by Kim and Jongman (1996). The authors investigated Coda Neutralization in Korean, whereby syllable final coronal obstruents (e.g. /t, tʰ, s/) are all phonetically realized as [t]. Korean vowels preceding non-final heterosyllabic fricatives are longer than when preceding non-final heterosyllabic oral stops. The authors used this distributional fact to measure vowel and closure durations of word-final VC sequences to determine whether the speech signal contained any acoustic cues to the underlying manner distinction. There was no indication of IN, neither numerically nor inferentially. Additionally, in a perception experiment, listeners were not able to distinguish underlying manner of articulation distinctions in the neutralizing context better than chance. The lack of any numerical evidence and the incapability of perceiving the contrast in neutralizing position is strong evidence for a complete neutralization. Since neutralized forms are morphological related (thus semantically and formally similar) to non neutralizing forms (cf. 1, taken from Kim & Jongman 1996: 298), Co-A makes the wrong prediction.

(1) /kəs/ ‘thing’ /kəs + i/ [kə.si] (with subject marker /i/)
/kəs + kwa/ [kət,k'wa] (with marker /kwa/ ‘and’)

However, as suggested above, IN is highly dependent on properties of the mental lexicon, therefore the absence of IN could be due to particular properties of the Korean lexicon. For example, the morphological verb paradigm in Korean is quite complex, consisting of voice suffixes, honorific suffixes, tense suffixes, “humble” suffixes and inflectional endings. The attachment of those suffixes leads sometimes to neutralization of stem final sounds, sometimes not. So when a target form is activated (e.g. /bʌs/ ‘undress’), many morphological related forms are co-activated. These lexical neighbors
contain and activate both phonological representations containing the non neutralized forms (e.g. /bʌs + əs/ [batgi] ‘undress.PASSIVE’) and the neutralized forms (e.g. /bʌs + əs/ [bʌsə] ‘undress.PAST’). The effect of co-activation could get lost either due to the large paradigm and therefore widely distributed co-activation, or due to the balanced proportion of neutralized and non-neutralized forms. Korean has a complex network of interconnected forms. Particular characteristics of lexical organisation might be inhibiting a co-activation process which produces IN. This, however, remains a speculation until further research has been conducted. The crucial point is, that the absence of IN is not necessarily a knock-out argument for the proposed account. The co-activation hypothesis has to be investigated carefully in light of the lexical organisation of a particular language. In this context, following dimensions of lexical organisation might be relevant: token frequency and type frequency of lexical neighbors, proportion of neutralized vs. non-neutralized neighbors, and similarity of neighbors to neutralizing target form.

For example, consider German minimal pairs used in the IN literature (those that are uncontroversial): Rad ‘wheel’ vs. Rat ‘council’; Tod ‘death’ vs. tot ‘dead’; Bund ‘association’ vs. bunt ‘colorful’; Bad ‘bath’ vs. bat ‘begged’. All of them are frequently used words in German. Using the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995), I looked for the most frequent morphological neighbors of the devoiced forms. For Rad (abs. frequency =151), the most frequent neighbor is the plural form Räder (=69) followed by the dative plural form Rädern (=27); for Tod (=712), it is the plural form Tode (=238), followed by the genitive form Todes (=66); for Bund (=177), beside the composita Bundesrepublik ‘federal republic’ and Bundesgebiet ‘federal territory’ (=2058 and 161, respectively), it is the genitive form Bundes (=92); for Bad (=821), beside the composita Badezimmer ‘bathroom’ and Badewanne ‘bathtub’ (=52 and 40, respectively), it is the verb baden ‘to bath’ (=29). In sum, all of the neutralized forms have highly frequent morphological neighbors with the corresponding non-neutralized variant. These non-target representations are predicted to be co-activated by the target representation resulting in IN effects. In turn, a language exhibiting lexical patterns in which target forms co-activate highly frequent non-target representations supporting the neutralized variant, might not show any IN effects.
Again, if I am on the right track, the co-activation hypothesis has to be investigated carefully in light of the lexical organisation of the language under investigation.
Concluding remarks

“The final devoicing rule is still a major issue in the phonetics-phonology interface.”
Fuchs (2005: 25)

A substantial number of experiments over the last three decades have reported on Incomplete Neutralizations of contrasts. Production studies showed subtle acoustic differences between two underlyingly different segments in a phonologically neutralizing context. These findings posed a problem for traditional feed forward models of phonological/phonetic encoding and led scholars to rethink phonological theory. Port and Crawford have recognized this and have claimed that IN findings “pose a threat to phonological theory” (1989: 257).

This paper reviewed the empirical evidence for IN, in particular IN of final devoicing. Based on detailed acoustic studies, it has been argued that there are small acoustic and articulatory differences between words such as Tod and tot. Moreover, listeners are sensitive to IN, since they can tell apart “devoiced” stops from voiceless ones with above-chance accuracy. It was argued that proposed generative phonological explanations for this phenomenon are post hoc explanations that do not serve any predictive purpose. The difficulty in incorporating IN into traditional models of phonological encoding might have led many scholars to neglect IN as a genuine linguistic phenomenon and claimed that IN was an artifact of the experimental design. These critiques have been tackled recently by Röttger et al. (2011). According to them IN has to be considered an effect unlikely to be only induced by methodological errors.
Since the IN effect seems to be robustly reproducible, alternative explanations must be considered. Within exemplar-based models of lexical organisation a possible explanation for IN was discussed: within a rich lexicon full of detailed interconnected information, IN is explained as an artifact of lexical co-activation. The mental lexicon is hypothesized to contain completely inflected forms. In German these inflected forms generally have obstruents in non-neutralizing positions (e.g. [ʁæ:ɖɛ] Räder ‘wheels’). Co-activation of these inflected morphological neighbors could therefore influence the production of corresponding neutralized forms such as [ʁaɛt] (Rad ‘wheel’), thus resulting in a phonetic difference between Rat and Rad (the IN effect). This account generated testable hypotheses which were validated empirically in the work at hand. The study investigated one acoustic parameter reported to be a robust IN cue across languages and across studies within German: duration of the vowel preceding the word-final stop. To this end, subjects had to produce “neutralized” singular forms such as /ɡɔp/ from auditory presented voicing-specified plural forms (/ɡɔpʰe/, /ɡoːbe/). The voicing contrast was kept intact with respect to voice onset time, closure duration and voicing, but the duration of the vowel preceding the stop was manipulated such that vowels preceding underlying voiced stops were of one of four lengths relative to the length of the voiceless stops: 32% longer (ENHANCED), 16% longer (REGULAR), the same length (NONE) or 16% shorter (REVERSED). The subjects’ productions, first, showed a significant overall IN effect, i.e., longer vowel duration in case of underlingly voiced as compared to underlingly voiceless consonants. Additionally, the size of this effect was modulated by the durational manipulation of the stimuli indicating that phonetic details of the paradigmatic neighbors as well as abstract prelexical categories influence the productions of derived forms.

These results are discussed regarding the representational status of IN. Here, it has been argued, that the difference of underlying voiced and voiceless stops in neutralized position (IN effect) must be somehow represented. If lexical co-activation causes Rat and Rad to surface acoustically different, a listener will store these exemplars topologically dispersed. This might result in exemplar clouds of two categories (underlying voiced and voiceless stops) which overlap substantially but the category membership remains distinct. Based on this assumption listeners are able to match the just perceived speech token with the averaged mean of the category exemplar cloud. So
listeners are able to discriminate underlying voiced from voiceless forms better than chance. Because of highly overlapping representations, the judgment accuracies are very low.

Finally, it has been argued that the proposed co-activation hypothesis has to be investigated carefully in light of the lexical organisation of a particular language. In this context, the following dimensions of lexical organisation have been suggested to play a role: token frequency and type frequency of lexical neighbors, proportion of neutralized vs. non-neutralized neighbors, and similarity of neighbors to neutralizing target form.

For future research, one should replicate the present results with a more balanced manipulation basis and a between-subject designs as suggested in §4.2. Moreover, since these findings have to be generalizable to the actual lexicon, experiments with real words have to be considered. Cross-linguistic research is desirable, too, to shed some light on the relationship of the properties of the lexicon and the manifestation of Incomplete Neutralization as discussed in §5.3.

Finally, I would like to relate the present account of IN to its functional relevance. An acoustic difference of several milliseconds elicited in a formal setting of an experiment, which, moreover, is barely identifiable in a forced choice experiment might be under the threshold of psycho-acoustic processing and thus negligible for real life communication. The present findings are agnostic about the functional relevance. However, from a functionalist perspective, an account of IN that is based on interference from paradigmatic representations is more attractive than an account based on phonetic or phonological rules and/or representations that are extracted from auditory information. Under a co-activation account, speakers would not need to extract any subtle contrast from the signal, at least as long as they do extract the contrast between the corresponding paradigmatic neighbors. The acquisition of the contrast within the paradigm would automatically cause interference through co-activation resulting in Incomplete Neutralisation.

Manaster Ramer (1996: 487) uses the Incomplete Neutralization debate as a call for phonologists and phoneticians to collaborate more with each other. He points out that
phoneticians need phonologists in order to design experiments that are phonologically meaningful, and phonologists need phoneticians to test claims of phonological theory, or to see whether purported problems to phonological theory are actually problematic. In Manaster Ramer’s words (ibid. 487), “Phonologists cannot afford to be neutral” with respect to Incomplete Neutralization. I have shown that the phenomenon can be seen in a different light if psycholinguistic and cognitive evidence is taken into account. So, not only do phonologists and phoneticians need to collaborate with each other, but both need to look more at work from other disciplines to gain new perspectives of old phenomena.

It is still an ongoing debate, how to marry off low-dimensional, discrete, and invariant aspects of our cognition (in this case, the representation of a contrast or the lack of it) and high-dimensional, continuous, and variant aspects of our performance (in this case the phonetic manifestation of a contrast). This phonology-phonetics relation is the central theme of the laboratory phonology research community (e.g. Beckman & Kingston, 1990; Ohala, 1990) Incomplete Neutralization lies obviously at the heart of that relation. Fuchs (2005: 172) thus rightly points out that IN “is still a major issue in the phonetics-phonology interface.”

The work in hand has tried to shed light on a less understood phenomenon in phonology by experimental work, thus trying to marriage discrete symbols of phonological theory and gradient performance observed in laboratory. However, the proposed account of the investigated phenomenon is ultimately psychological, stressing the importance of cognitive considerations in linguistic theory.


categories are necessary for coping with speaker-related variation. In C. Fougeron,
B. Kühnert, M. D’Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 91–
111). Berlin: de Gruyter.
Evidence from the comprehension of noise-vocoded sentences. *Journal of
Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical
garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of
Experimental Psychology: Human Perception and Performance*, 28, 218–244.
de Jong, K. J. (2011). Flapping in American English. In M. van Oostendorp, C. J. Ewen,
E. Hume, & K. Rice (Eds.), Blackwell Companion to Phonology (pp. 2711–2729).
B. Brown, K. Walters, & J. Baugh (Eds.), *Linguistic change and contact:
Proceedings of the 16th. Annual Conference on New Ways of Analyzing Variation
in Language* (pp. 84–92). Austin: University of Texas.
Dinnsen, D. A., & Garcia-Zamor, M. (1971). The three degrees of vowel duration in
implementation and individual differences. *Journal of Phonetics*, 12, 49–60.
Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech
Ernestus, M., & Baayen, R. H. (2006). The functionality of incomplete neutralization in
Best (Eds.), *Laboratory Phonology 8* (pp. 27–49). Berlin: de Gruyter.


Kohler, K. J. (2007). Beyond Laboratory Phonology. The phonetics of speech communication, In M.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental Approaches to Phonology* (pp. 41–53). Oxford: Oxford University Press.


### Appendix B: critical stimuli

<table>
<thead>
<tr>
<th>[+voice]</th>
<th>[-voice]</th>
<th>PLACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bauge</td>
<td>[baʊɡə]</td>
<td>Bauke</td>
</tr>
<tr>
<td>Bege</td>
<td>[beɡə]</td>
<td>Beke</td>
</tr>
<tr>
<td>Blebe</td>
<td>[bleːba]</td>
<td>Blepe</td>
</tr>
<tr>
<td>Bloge</td>
<td>[bloːda]</td>
<td>Bloke</td>
</tr>
<tr>
<td>Dage</td>
<td>[daːɡə]</td>
<td>Dake</td>
</tr>
<tr>
<td>Dabe</td>
<td>[daːbə]</td>
<td>Dape</td>
</tr>
<tr>
<td>Diege</td>
<td>[diːɡə]</td>
<td>Dieke</td>
</tr>
<tr>
<td>Dobe</td>
<td>[doːbə]</td>
<td>Dope</td>
</tr>
<tr>
<td>Drube</td>
<td>[dʁuːbə]</td>
<td>Drupe</td>
</tr>
<tr>
<td>Duge</td>
<td>[duːɡə]</td>
<td>Duke</td>
</tr>
<tr>
<td>Fage</td>
<td>[faːɡə]</td>
<td>Fake</td>
</tr>
<tr>
<td>Faube</td>
<td>[faʊbə]</td>
<td>Faupe</td>
</tr>
<tr>
<td>Flabe</td>
<td>[flaːbə]</td>
<td>Flape</td>
</tr>
<tr>
<td>Flebe</td>
<td>[fleːbə]</td>
<td>Flepe</td>
</tr>
<tr>
<td>Frebe</td>
<td>[fɾeːbə]</td>
<td>Frepe</td>
</tr>
<tr>
<td>Froge</td>
<td>[fɾoːɡə]</td>
<td>Froke</td>
</tr>
<tr>
<td>Frobe</td>
<td>[fɾoːbə]</td>
<td>Frobe</td>
</tr>
<tr>
<td>Frube</td>
<td>[fɾuːbə]</td>
<td>Frupe</td>
</tr>
<tr>
<td>Gaube</td>
<td>[ɡaʊbə]</td>
<td>Gaupe</td>
</tr>
<tr>
<td>Gage</td>
<td>[ɡaʊɡə]</td>
<td>Gauke</td>
</tr>
<tr>
<td>Glege</td>
<td>[ɡleːɡə]</td>
<td>Gleke</td>
</tr>
<tr>
<td>Gliebe</td>
<td>[ɡliːbə]</td>
<td>Gliepe</td>
</tr>
<tr>
<td>Gobe</td>
<td>[ɡoːbə]</td>
<td>Gope</td>
</tr>
<tr>
<td>Griebe</td>
<td>[ɡʁiːbə]</td>
<td>Griepe</td>
</tr>
<tr>
<td>Hege</td>
<td>[heːgə]</td>
<td>Heke</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Klabe</td>
<td>[klæbə]</td>
<td>Klapo</td>
</tr>
<tr>
<td>Krobe</td>
<td>[kʁoːbə]</td>
<td>Krope</td>
</tr>
<tr>
<td>Miebe</td>
<td>[miːbə]</td>
<td>Miepe</td>
</tr>
<tr>
<td>Naube</td>
<td>[naʊ̯bə]</td>
<td>Naupe</td>
</tr>
<tr>
<td>Nauge</td>
<td>[naʊ̯gə]</td>
<td>Nauke</td>
</tr>
<tr>
<td>Priege</td>
<td>[pʁiːgə]</td>
<td>Priepe</td>
</tr>
<tr>
<td>Pruge</td>
<td>[pʁʊːgə]</td>
<td>Pruke</td>
</tr>
<tr>
<td>Roge</td>
<td>[ʁoːgə]</td>
<td>Roke</td>
</tr>
<tr>
<td>Schlabe</td>
<td>[ʃlaːbə]</td>
<td>Schlape</td>
</tr>
<tr>
<td>Schmaube</td>
<td>[ʃmaʊ̯bə]</td>
<td>Schmaupe</td>
</tr>
<tr>
<td>Schriege</td>
<td>[ʃʁiːgə]</td>
<td>Schrieke</td>
</tr>
<tr>
<td>Schuge</td>
<td>[ʃʊːgə]</td>
<td>Schuke</td>
</tr>
<tr>
<td>Spage</td>
<td>[ʃpaːgə]</td>
<td>Spake</td>
</tr>
<tr>
<td>Stauge</td>
<td>[ʃtaʊ̯gə]</td>
<td>Stauke</td>
</tr>
<tr>
<td>Strege</td>
<td>[ʃtʁeːgə]</td>
<td>Streke</td>
</tr>
<tr>
<td>Sube</td>
<td>[zʊːbə]</td>
<td>Supe</td>
</tr>
<tr>
<td>Triege</td>
<td>[tʁiːgə]</td>
<td>Trieke</td>
</tr>
<tr>
<td>Wiebe</td>
<td>[viːbə]</td>
<td>Wiepe</td>
</tr>
<tr>
<td>Wube</td>
<td>[vʊːbə]</td>
<td>Wupe</td>
</tr>
<tr>
<td>Zebe</td>
<td>[tseːbə]</td>
<td>Zepe</td>
</tr>
</tbody>
</table>
### Appendix B: demonstration stimuli

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blame</td>
<td>[blaːmə]</td>
</tr>
<tr>
<td>Bome</td>
<td>[boːmə]</td>
</tr>
<tr>
<td>Daule</td>
<td>[daʊ̯lə]</td>
</tr>
<tr>
<td>Drale</td>
<td>[dʁaːlə]</td>
</tr>
<tr>
<td>Fene</td>
<td>[feːnə]</td>
</tr>
<tr>
<td>Fume</td>
<td>[fuːmə]</td>
</tr>
<tr>
<td>Gale</td>
<td>[gaːlə]</td>
</tr>
<tr>
<td>Grieme</td>
<td>[ɡʁiːmə]</td>
</tr>
<tr>
<td>Kome</td>
<td>[kʰoːnə]</td>
</tr>
<tr>
<td>Kriene</td>
<td>[kʁiːnə]</td>
</tr>
<tr>
<td>Mene</td>
<td>[meːnə]</td>
</tr>
<tr>
<td>Schlune</td>
<td>[ʃluːnə]</td>
</tr>
<tr>
<td>Spieme</td>
<td>[ʃpiːmə]</td>
</tr>
<tr>
<td>Tome</td>
<td>[tɔmə]</td>
</tr>
<tr>
<td>Waule</td>
<td>[vaʊ̯lə]</td>
</tr>
<tr>
<td>Zaule</td>
<td>[tsaʊ̯lə]</td>
</tr>
</tbody>
</table>
### Appendix C: training stimuli

<table>
<thead>
<tr>
<th>Term</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flene</td>
<td>[fleːna]</td>
</tr>
<tr>
<td>Game</td>
<td>[gaːmə]</td>
</tr>
<tr>
<td>Hiene</td>
<td>[hiːnə]</td>
</tr>
<tr>
<td>Schlume</td>
<td>[ʃluːmə]</td>
</tr>
<tr>
<td>Strame</td>
<td>[ʃtʁaːmə]</td>
</tr>
<tr>
<td>Traule</td>
<td>[traʊ̯lə]</td>
</tr>
<tr>
<td>Wole</td>
<td>[voːlə]</td>
</tr>
<tr>
<td>Wule</td>
<td>[vuːlə]</td>
</tr>
</tbody>
</table>
## Appendix D

|                     | Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|---------------------|----------|----------|------------|------------|-------|----------|
| (Intercept)         | 0.1506   | 0.1503   | 0.1418     | 0.1585     | 0.0001| 0.0000   |
| voicing_plus        | 0.0043   | 0.0043   | 0.0030     | 0.0056     | **0.0001**| **0.0000**|
| condition1          | -0.0042  | -0.0042  | -0.0096    | 0.0013     | 0.1220| 0.1857   |
| condition2          | -0.0019  | -0.0018  | -0.0076    | 0.0043     | 0.5350| 0.5865   |
| condition3          | 0.0011   | -0.0011  | -0.0079    | 0.0057     | 0.7436| 0.7775   |
| repetition1         | -0.0020  | -0.0020  | -0.0034    | -0.0006    | **0.0066**| **0.0058**|
| repetition2         | -0.0018  | -0.0018  | -0.0034    | -0.0002    | **0.0316**| **0.0261**|
| place_velar         | 0.0058   | 0.0062   | 0.0025     | 0.0102     | **0.0016**| **0.0076**|
| vowel_au            | -0.0051  | -0.0051  | -0.0129    | 0.0025     | 0.1980| 0.2766   |
| vowel_e             | -0.0351  | -0.0351  | -0.0424    | -0.0268    | **0.0001**| **0.0000**|
| vowel_i             | -0.0485  | -0.0485  | -0.0562    | -0.0409    | **0.0001**| **0.0000**|
| vowel_o             | -0.0335  | -0.0334  | -0.0415    | -0.0254    | **0.0001**| **0.0000**|
| vowel_u             | -0.0526  | -0.0525  | -0.0606    | -0.0452    | **0.0001**| **0.0000**|
| voicing_plus:condition1 | 0.0024  | 0.0024   | -0.0001    | 0.0048     | 0.0562| 0.0552   |
| voicing_plus:condition2 | 0.0019  | 0.0020   | -0.0017    | 0.0058     | 0.3074| 0.3060   |
| voicing_plus:condition3 | 0.0014  | 0.0015   | -0.0022    | 0.0053     | 0.4392| 0.4372   |

Figure A: Fixed effects of *Omnibus Model*; MCMC simulation results, 10,000 runs.
|                      | Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|----------------------|----------|----------|------------|------------|-------|---------|
| (Intercept)          | 0.1451   | 0.1449   | 0.1340     | 0.1558     | 0.0001| 0.0000  |
| voicing_plus         | 0.0060   | 0.0060   | 0.0042     | 0.0079     | 0.0001| 0.0000  |
| condition4           | 0.0054   | 0.0055   | -0.0011    | 0.0127     | 0.1056| 0.1638  |
| repetition1          | -0.0038  | -0.0039  | -0.0066    | -0.0010    | 0.0076| 0.0072  |
| repetition2          | -0.0043  | -0.0043  | -0.0076    | -0.0012    | 0.0104| 0.0076  |
| place_velar          | 0.0043   | 0.0044   | -0.0006    | 0.0099     | 0.1004| 0.1526  |
| vowel_au             | -0.0035  | -0.0034  | -0.0148    | 0.0082     | 0.5446| 0.5980  |
| vowel_e              | -0.0344  | -0.0343  | -0.0462    | -0.0228    | 0.0001| 0.0000  |
| vowel_i              | -0.0424  | -0.0424  | -0.0541    | -0.0311    | 0.0001| 0.0000  |
| vowel_o              | -0.0278  | -0.0277  | -0.0395    | -0.0168    | 0.0001| 0.0000  |
| vowel_u              | -0.0482  | -0.0482  | -0.0601    | 0.0371     | 0.0001| 0.0000  |
| voicing_plus:condition4 | -0.0038 | -0.0038 | -0.0065 | -0.0012 | 0.0040 | 0.0043 |
| voicing_plus:repetition1 | 0.0011 | 0.0012 | -0.0027 | 0.0052 | 0.5616 | 0.5784 |
| voicing_plus:repetition2 | 0.0046 | 0.0046 | 0.0000 | 0.0090 | 0.0424 | 0.0427 |
| condition4:repetition1 | 0.0042 | 0.0042 | 0.0002 | 0.0082 | 0.0396 | 0.0380 |
| condition4:repetition2 | 0.0034 | 0.0034 | -0.0009 | 0.0079 | 0.1394 | 0.1397 |
| voicing_plus:condition4:repetition1 | -0.0028 | -0.0028 | -0.0087 | 0.0027 | 0.3210 | 0.3214 |
| voicing_plus:condition4:repetition2 | -0.0058 | -0.0058 | -0.0121 | 0.0004 | 0.0682 | 0.0695 |

Figure B: Fixed effects of Subset Model; MCMC simulation results, 10,000 runs.